On the Heat Channel and Its Capacity*

Edwin Hammerich
Ministry of Defence
Kulmbacher Str. 58–60, D-95030 Hof, Germany
Email: edwin.hammerich@ieee.org

Abstract

The heat channel is defined by an analog filter and a subsequent measurement of the filter output signal perturbed by additive white Gaussian noise. The filter is related to the heat kernel of the quantum mechanical harmonic oscillator, so the name of the channel. The channel is modeled as an infinite-dimensional vector Gaussian channel and the capacity in terms of average energy of the input signal is derived along with a method of capacity achieving signaling for the continuous-time channel. Then, a characterization of the capacity by water-filling in the time-frequency plane is stated and proved. We compare our findings with a classical capacity result of Gallager. A related problem in rate distortion theory is investigated to some extent. Finally, a second formula for the capacity of the heat channel based on average energy of the measured perturbed filter output signal is derived. The result is interpreted in context of estimation theory and a parallel to the I-MMSE relationship due to Guo et al. is presented connecting the capacity of the heat channel with an estimation error for the output.

1 Introduction

The conduction of heat in solid bodies was mathematically described and solved by Joseph Fourier in his fundamental 1822 treatise *Théorie analytique de la chaleur* [1]. In one dimension, *e.g.*, in case of a heat-conducting insulated wire, his description results in the partial differential equation (the heat equation)

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2},$$

in which $u = u(x,t)$ is temperature at time $t \geq 0$ at any point x and k is a positive constant depending on the material. Given the initial temperature

*The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Seoul, Korea, June 28–July 3, 2009.

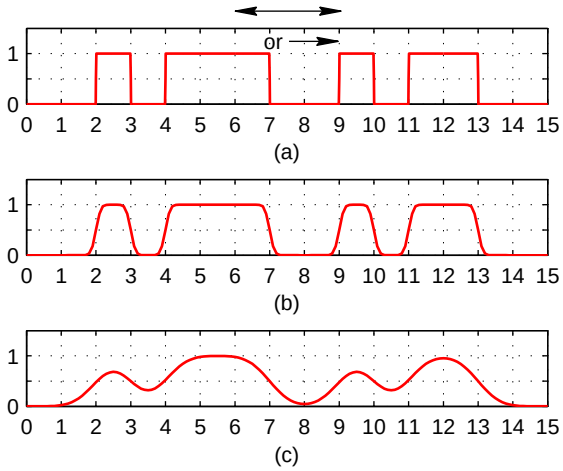

Figure 1: Spreading of heat/Signal spread because of dispersion in an optical fiber (attenuation not regarded). (a) Initial temperature distribution/Fiber input signal, $f(x)$. (b) Temperature, shortly after/Output signal, short fiber ($\beta = 10$). (c) Temperature, later time/Output signal, longer fiber ($\beta = 2$). Arrows indicate direction of spatial propagation of heat ($\leftrightarrow$) or optical intensity ($\rightarrow$).

distribution $f(x)$ for a wire of infinite length, Fourier's solution of the heat equation is

$$u(x,t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4kt}} f(y)\, dy. \tag{1}$$

Since, in general, for any time $t > 0$ the inversion of the integral transform appearing in (1) is unfeasible in practice [2], we observe an unavoidable loss of "information" (in a preliminary, informal sense). In Fig. 1, several temperature distributions $u(x,t)$ are depicted showing how the initial one is gradually smeared out by the propagation of heat.

A similar situation, in principle known since the earliest days of cable communication [3], arises in fiber optics. In a transmission through a (single-mode) optical fiber, signals experience besides attenuation a spread over time due to (chromatic) dispersion (see, *e.g.*, [4], [5]). A frequently used model for dispersion in an optical fiber [4] is a linear time-invariant (LTI) filter with impulse response (*cf.* Fig. 2)

$$h_1(t) = \frac{1}{\sqrt{2\pi}(1/\beta)} e^{-\frac{1}{2}\frac{t^2}{(1/\beta)^2}}, \tag{2}$$

i.e., a Gaussian filter with the standard deviation $1/\beta$, $\beta \in (0, \infty)$ characterizing dispersion (the parametrization is chosen to fit later notation). The fiber

input/output relation is now given by the convolution integral

$$u(t) = a \int_{-\infty}^{\infty} h_1(t - t') f(t')\, dt', \tag{3}$$

where $f \in L^2(\mathbb{R})$ is the finite-energy input signal and $u(t)$ the output, the constant factor $a \in (0, 1]$ representing attenuation. Except for the factor a and change of physical dimension from position (variable x) to time (variable t), we now observe perfect analogy between Eqs. (1) and (3). As a consequence, in Fig. 1 also the degradation of an optical signal—initially a sequence of bits (here, binary symbols) obtained by intensity modulation and on-off keying—by dispersion in an optical fiber is displayed. Obviously, dispersion limits the information throughput of an optical fiber because of intersymbol interference (ISI). A maximum attainable bit rate can be estimated by considering as fiber input a sequence of unit impulses separated by time intervals of duration T, the output then being a sequence of Gaussian pulses of standard deviation $\Delta\tau = 1/\beta$. In order to cope with ISI, a popular criterion is $\Delta\tau \leq T/4$ [4] resulting in our case for the bit rate $R_b = 1/T$ (in binary symbols per second) in the estimate

$$R_b \leq \frac{\beta}{4}.$$

If the fiber output signal is corrupted by noise, this rule of thumb becomes questionable. In case of additive white Gaussian noise (AWGN), typically caused by an optical amplifier [4], [6], we arrive at a continuous-time (or waveform) channel following the model in Gallager's 1968 book [7]; see Fig. 2(a). Here, of course, we supposed that, *e.g.*, the power spectral density (PSD) of the AWGN is independent of the input (see the recent paper [8] for an opposite situation), and the input power is not so high that arising nonlinearities [4] would destroy the linear model (3); we refer to [5] for a variety of possible other perturbations that limit the capacity of optical fiber communication systems. Gallager's waveform channel, consisting of an LTI filter and *subsequent* AWGN (in [7] even nonwhite Gaussian noise is considered), yet may be viewed as a generalization of the bandlimited Gaussian channel of Shannon [9], [10], see Fig. 2(b), because in case of ideal low-pass filters and AWGN the two channel models are equivalent. The capacity of the Gallager channel is given in the result [7, Theorem 8.5.1], by many referenced as "the" solution to the capacity problem [11]. When applied to the above case of a Gaussian LTI filter and AWGN, several questions arise in connection with Gallager's result. First, capacity is given in parametric form and a closed-form expression remains a challenge. Secondly, there is no apparent method for capacity achieving signaling in [7, Chapter 8]. Finally, as will become evident in the present paper, the outcome of Gallager's capacity formula would considerably underestimate the attainable capacity (or spectral efficiency) of a communication system containing the model of a Gaussian filter and subsequent AWGN (possibly among other features); actually, addition of certain system components as indicated in Fig. 2(a) would lead to a compound channel of significantly higher spectral efficiency.

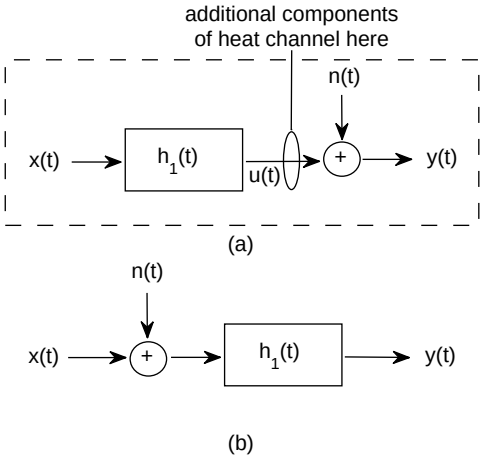

Figure 2: Continuous-time (or waveform) channels. (a) Gallager model [7, Figure 8.4.1] (in dashed box); LTI filter with impulse response $h_1(t) \in L^2(\mathbb{R})$. (b) Bandlimited Gaussian Channel [9], [10]; here, $h_1(t)$ is the impulse response of an ideal low-pass filter. Input signal $x(t)$ always power-limited, and noise waveform $n(t)$ realization of white Gaussian noise.

In this paper, we investigate the linear time-varying (LTV) filter (or operator) $\boldsymbol{P}_\delta^{(\gamma)} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ given by

$$(\boldsymbol{P}_\delta^{(\gamma)} f)(t) = e^{-\frac{t^2}{2\alpha^2}} \frac{\beta}{\sqrt{2\pi \cosh \delta}} \int_{-\infty}^{\infty} \exp\left[-\frac{\beta^2}{2}\left(\frac{t}{\cosh \delta} - t'\right)^2\right] f(t')\, dt', \quad (4)$$

where t stands for time and α, β are any positive numbers satisfying $\alpha\beta > 1$; the positive parameters γ, δ are defined by $\gamma^2 = \alpha/\beta$, $\coth \delta = \alpha\beta$. This operator is an instance of a time-frequency localization operator used in signal analysis for the approximate concentration of a signal in both time and frequency [12]. Operators based on radial Gaussian weights on the time-frequency plane were introduced by Daubechies in the seminal paper [13]; a generalization to arbitrary Gaussian weights leads to the above operator [14]. The Fourier transform of the filter output signal $g = \boldsymbol{P}_\delta^{(\gamma)} f$ is

$$\hat{g}(\omega) = e^{-\frac{\omega^2}{2\beta^2}} \frac{\alpha}{\sqrt{2\pi \cosh \delta}} \int_{-\infty}^{\infty} \exp\left[-\frac{\alpha^2}{2}\left(\frac{\omega}{\cosh \delta} - \omega'\right)^2\right] \hat{f}(\omega')\, d\omega', \quad (5)$$

where ω is angular frequency and the convention $\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-it\omega} f(t)\, dt$ for the Fourier transform has been used [12]. The condition $\alpha\beta > 1$ is now seen as imposed by the uncertainty principle of communications [15]. Interestingly enough, the kernel of operator $\boldsymbol{P}_\delta^{(\gamma)}$ coincides with the heat kernel [16] of the

quantum mechanical harmonic oscillator.[1] Because of the Gaussian prefactor on the right-hand side (RHS) of Eq. (5), the Fourier transform $\hat{g}(\omega)$ of the filter output decays exponentially outside of the interval $[-\pi\beta, \pi\beta]$ (provided that the energy of the input f is not too large). Thus, g may be considered an approximately bandlimited signal of approximate bandwidth $W = \beta/2$ in positive frequencies measured in hertz.

If $u(t)$ is the output of the LTI filter (3) (where $a = 1$) with impulse response (2) upon input f, then the output g of the LTV filter (4) may be written as

$$g(t) = e^{-\frac{t^2}{2\alpha^2}} \cdot (\cosh\delta)^{-\frac{1}{2}} u(t/\cosh\delta). \tag{6}$$

Thus, $g(t)$ is just the dilated LTI filter output $u(t)$ multiplied with a Gaussian time window; notice that the dilation factor, $\cosh\delta$, is close to one when $\alpha\beta$ is large. In Fig. 2(a), those two operations are quoted as additional components of the heat channel, the latter meaning the continuous-time channel formed by the LTV filter (4) and subsequent AWGN. A precise formal definition of the heat channel (whose name is chosen after the aforementioned heat kernel) will be given in Section 2.

The goal of the present paper is to quantify the capacity of the heat channel in various ways, to relate the results to other work, and to discuss the implications.

1.1 Related Work

The computation of the capacity of time-invariant waveform channels, pioneered by Holsinger [17], was put on a firm ground by Gallager [7] who gave a rigorous proof for the water-filling characterization in [7, Theorem 8.5.1]. The overall approach rests on orthogonal expansion of the channel input, resulting in a discretization of the continuous-time channel in form of parallel Gaussian channels whose capacity is known; the corresponding discrete water-filling formulas are also to be found in [7] (where they are attributed to Ebert [18]). The result is then transposed to the frequency domain by means of a specific Szegö theorem (whose proof in [7] is based on previous Szegö type results in [19], [20]). The quest for water-filling characterizations for the capacity of time-varying channels in the time-frequency plane is now an active area of research; see [21], [22], [23] to cite only a few. A water-filling characterization for time-varying channels that are periodic in time has recently been given by Jung [21]. Our Szegö theorem (Theorem 6), based on a Weyl symbol connected with the LTV filter, is perhaps closest to the one proved in [24]. The use of Weyl symbols (and the related Wigner-Ville spectrum) has a long history in communications; see, *e.g.*, [25], [26], [27]. Concerning reverse water-filling representations for the rate distortion function of stationary Gaussian processes in the frequency domain we refer to Berger [28] (his approach is also based on [19], [20]). The "C-LLSE

[1] In [16, p. 114], the heat kernel of the one dimensional quantum mechanical harmonic oscillator with Hamiltonian $H = -\frac{d^2}{dx^2} + a^2x^2$ takes the form of the kernel of operator (4) after the substitution $2at = \delta$, $a = \gamma^{-2}$.

formula" in Section 5.2 and the ensuing discussion draws inspiration from Guo *et al.* [29]. The I-MMSE relationship in connection with SNR-dependent input signals at low SNR is treated in the recent paper [30].

1.2 Contribution

We derive a closed-form expression for the capacity of the heat channel in terms of average energy of the input (Theorem 1) and infer a similar expression in terms of average input power. Moreover, we present a method of capacity achieving signaling—even in form of short pulses when the dispersion parameter β is large enough. The water-filling characterization for the capacity of the heat channel in the time-frequency plane (Theorem 2) is the first clear-cut, full-fledged example of its kind for a time-varying channel. We find the surprising fact that the spectral efficiency of the heat channel in the limiting case $\alpha \to \infty$ (and SNR not too small) is significantly higher than that of the corresponding Gallager channel, *i.e.*, the time-invariant waveform channel with Gaussian LTI filter and subsequent AWGN. The rate distortion function in closed form (Theorem 3) is one of the rare examples of such a representation in rate distortion theory; the reverse water-filling representation in the time-frequency plane (Theorem 4) is a novel result for a nonstationary Gaussian process. The (reverse) water-filling formulas are based on an new, specific Szegö theorem (Theorem 6) for which we give a straightforward, essentially self-contained proof (the only external ingredient is the "trace rule" from [24]). We supplement the I-MMSE relationship discovered by Guo *et al.* [29] by a new "C-LLSE formula" (Proposition 1) that connects, in the specific context of the heat channel, increase of channel capacity with an estimation error for the channel output (a setting where the channel input is necessarily SNR-dependent). The capacity results in this paper may serve as lower bounds for the attainable information throughput of single-mode optical fiber communication systems in the presence of chromatic dispersion and amplifier noise.

1.3 Organization

The remainder of the paper is organized as follows. In Section 2, the heat channel is defined. In Section 3, a closed formula for the capacity of the heat channel is given along with optimal signaling; the water-filling characterization of channel capacity in the time-frequency plane is also presented here. A related problem in rate distortion theory is investigated in Section 4. Section 5 deals with a second closed capacity formula for the heat channel and the relation to estimation theory. Finally, Section 6 concludes the paper.

2 The Heat Channel

The time-varying channel formed by the LTV filter (4) and subsequent AWGN, *cf.* Fig. 2(a), will now be reduced to an ordered set of parallel Gaussian channels

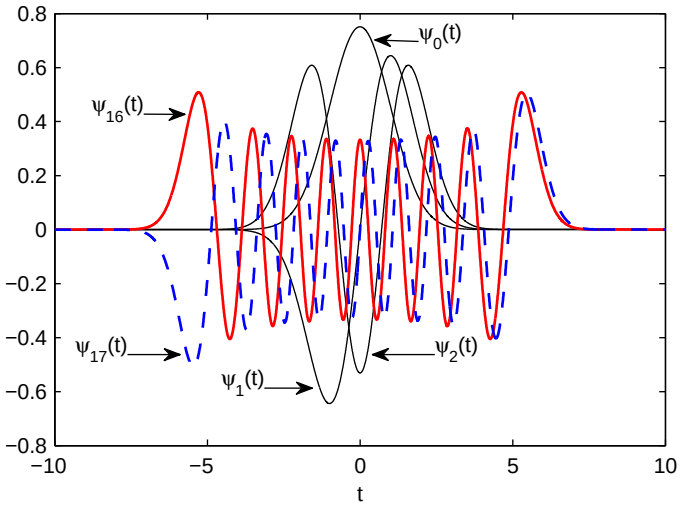

Figure 3: Hermite functions $\psi_0(t)$ (a normalized Gaussian function), $\psi_1(t)$, $\psi_2(t)$, $\psi_{16}(t)$ and $\psi_{17}(t)$. Strong decay in time is complemented by the same behaviour in the frequency domain since Hermite functions are eigenfunctions of the Fourier transform.

(or a vector Gaussian channel). We follow the overall approach of [7] where it was applied to time-invariant channels.

From now on we use the parameter

$$\rho = e^{-\delta},\ \delta = \operatorname{arccoth}(\alpha\beta) = \frac{1}{\alpha\beta} + \frac{1}{3(\alpha\beta)^3} + \dots . \tag{7}$$

Note that $\delta \sim \frac{1}{\alpha\beta}$, or $\delta(\alpha\beta) = \frac{1}{\alpha\beta} + o(\frac{1}{\alpha\beta})$, as $\alpha\beta \to \infty$.[2]

2.1 Diagonalization of the Filter

As shown in [13] in the radial case $\alpha = \beta$, *i.e.*, $\gamma = 1$ (see [14] for the general case $\gamma > 0$), the operator $\boldsymbol{P}_\delta^{(\gamma)}$ in (4) has eigenvalues $\rho^{k+\frac{1}{2}}$, $k = 0, 1, \dots$, with corresponding eigenfunctions

$$(D_\gamma \psi_k)(t) = \gamma^{-\frac{1}{2}} \psi_k(t/\gamma),$$

where $\psi_k(t) = (2^k k! \sqrt{\pi})^{-1/2} H_k(t) e^{-t^2/2}$ is the kth Hermite function, $H_k(t) = e^{t^2}(-d/dt)^k e^{-t^2}$ being the kth Hermite polynomial [31]. Since $\{D_\gamma \psi_k;\ k = 0, 1, \dots\}$ forms a complete orthonormal basis of $L^2(\mathbb{R})$, any function $f \in L^2(\mathbb{R})$ has an expansion $f(t) = \sum_{k=0}^{\infty} x_k\, (D_\gamma \psi_k)(t)$ where the coefficient sequence $x_0, x_1, \dots$ is an element of the space $\ell^2(\mathbb{N}_0)$ of square-summable complex sequences with index set $\mathbb{N}_0 = \{0, 1, \dots\}$. Hence for any filter input signal

[2] We use the standard Landau symbols $O(\cdot)$ ("big-O") and $o(\cdot)$ ("little-o").

$f \in L^2(\mathbb{R})$, the filter output signal has the representation

$$(\boldsymbol{P}_\delta^{(\gamma)} f)(t) = \sum_{k=0}^{\infty} \rho^{k+\frac{1}{2}} x_k \, (D_\gamma \psi_k)(t), \tag{8}$$

where $x_k = \langle f, D_\gamma \psi_k \rangle$, $\langle f_1, f_2 \rangle = \int_{-\infty}^{\infty} f_1(t) \overline{f_2(t)} \, dt$ denoting the inner product in $L^2(\mathbb{R})$. The new coefficient sequence is $(\rho^{k+\frac{1}{2}} x_k)_{k=0}^{\infty}$ and again an element of $\ell^2(\mathbb{N}_0)$. Thus, the filter $\boldsymbol{P}_\delta^{(\gamma)}$ is reduced to a diagonal linear transformation in $\ell^2(\mathbb{N}_0)$.

In Fig. 3, some Hermite functions are depicted; observe their strong decay in time (and frequency).

2.2 Channel Model

The noiseless filter output signal $g = \boldsymbol{P}_\delta^{(\gamma)} f \in L^2(\mathbb{R})$ is corrupted by AWGN of, say, two-sided PSD $N_0/2 = \theta^2$. Upon observation of the noisy filter output signal $\tilde{g}(t) = g(t) + n(t)$, where $n(t)$ is a realization of the noise, we want to reconstruct the filter input signal $f \in L^2(\mathbb{R})$ as accurately as possible. In the noiseless case, computation of the coefficients $y_k = \langle g, D_\gamma \psi_k \rangle = \rho^{k+\frac{1}{2}} x_k$ would allow to recover the coefficients $x_k, k = 0, 1, \ldots$, thus capturing f.

In the presence of noise, the best we can do is to measure the coefficients y_k by *optimal detection* (due to North [32]), for example by means of a matched filter [33]; in our case the LTI filter(s) with impulse response $h_k(t) = (D_\gamma \psi_k)(-t)$. When applied to the noisy signal $\tilde{g}$ as input, we get for the output at sampling instant $t_0 = 0$

$$\int_{-\infty}^{\infty} h_k(t_0 - t') \tilde{g}(t') \, dt' = \rho^{k+\frac{1}{2}} x_k + \int_{-\infty}^{\infty} h_k(-t') n(t') \, dt'.$$

From the theory of LTI filters we know (*cf.* [7, p. 365]) that the integral on the RHS evaluates to a realization n_k of a zero-mean Gaussian random variable N_k with the variance $\theta^2 \int_{-\infty}^{\infty} h_k^2(-t) \, dt$. Since any waveform $D_\gamma \psi_k$ has L^2 norm one, the variance of N_k is θ^2 and, thus, does not depend on k. Moreover, because of orthogonality of the waveforms, the random variables N_k are independent. Consequently, the detection errors n_k are realizations of independent identically distributed (i.i.d.) zero-mean Gaussian random variables $N_k \sim \mathcal{N}(0, \theta^2), k = 0, 1, \ldots$. Note that the noise PSD θ^2, measured in watts/Hz, has also the dimension of an energy.

Instead of y_k we now have obtained $\hat{y}_k = \rho^{k+\frac{1}{2}} x_k + n_k$. So we get the estimate $\hat{x}_k = \rho^{-k-\frac{1}{2}} \hat{y}_k = x_k + z_k$ for x_k, where z_k are realizations of independent Gaussian random variables $Z_k \sim \mathcal{N}(0, \theta^2 \rho^{-2k-1})$. Thus, we are led to the following definition.

Definition 1 *Let θ be any positive number and let $\rho \in (0, 1)$ be as in (7). Then the heat channel is the infinite-dimensional vector Gaussian channel*

$$Y_k = X_k + Z_k, \; Z_k \sim \mathcal{N}(0, \theta^2 \rho^{-2k-1}), \; k = 0, 1, \ldots, \tag{9}$$

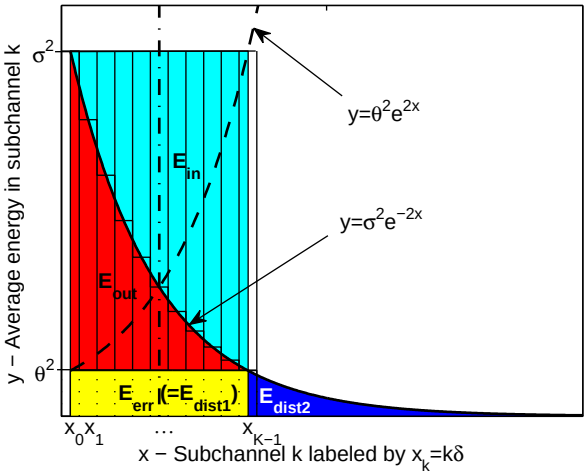

Figure 4: Balance of (average) energies around the heat channel: input/output energy ($E_{\rm in}/E_{\rm out}$), energy of measurement error ($E_{\rm err}$), and distortion ($E_{\rm dist1}+E_{\rm dist2}$). Subchannels displayed at distance $\delta\sim\frac{1}{\alpha\beta}$ apart as $\alpha\beta\to\infty$.

where the noise Z_k is assumed to be independent from subchannel to subchannel.

The extra factor ρ^{-1} in the noise variances is, of course, of no relevance; for the interpretation of θ^2 as entropy power of the measurement error we refer to [34].

$\boldsymbol{X}^K=(X_0,\dots,X_{K-1})^T$ will denote a K-dimensional column vector, $K\in\mathbb{N}$, of not necessarily independent random variables X_k. For any $S>0$, the vector Gaussian channel consisting of the first K subchannels of the heat channel has capacity [9], [10]

$$C^K(S)=\max_{\mathbb{E}\{\|\boldsymbol{X}^K\|^2\}\le S} I(\boldsymbol{X}^K;\boldsymbol{Y}^K), \tag{10}$$

where $I(\boldsymbol{X}^K;\boldsymbol{Y}^K)$ is the mutual information between input vector $\boldsymbol{X}^K$ and output vector $\boldsymbol{Y}^K$, $\boldsymbol{X}^K$ subject to the average energy constraint $\mathbb{E}\{\|\boldsymbol{X}^K\|^2\}=\sum_{k=0}^{K-1}\mathbb{E}X_k^2\le S$. The noise variances $\theta^2\rho^{-2k-1}$, $k=0,1,\dots,$ are monotonically increasing and unbounded. Consequently, by reason of the well-known water-filling argument (see, *e.g.*, [10]), for any fixed average input energy S the sequence of capacities $C^1(S),C^2(S),\dots$ eventually becomes constant. We define

$$C(S)=\lim_{K\to\infty} C^K(S) \tag{11}$$

as the capacity of the heat channel; $C(S)$ is measured in bits per channel use (or transmission). In Fig. 4, $E_{\rm in}=S$, the K subchannels to be read in reverse order (as will become clear in Section 3.1); further details of Fig. 4 will be described in the text.

2.3 Example: Dispersion and Amplifier Noise in Fiber Optics

When, as supposed here, in the fiber-optic transmission intensity modulation is used, real-valued input signals $f \in L^2(\mathbb{R})$ need to be replaced by waveforms $\tau_0 + f(t)$ where τ_0 is a fixed positive number chosen large enough so that the resulting signals are nonnegative with high probability [*cf.* Fig. 9(a), below]. We model dispersion by an LTI filter with impulse response (2). The dispersion parameter $\beta = \beta(L) \in (0, \infty)$ in (2) depends on fiber length L as well as the coefficient $a = a(L) \in (0, 1]$ in (3) representing attenuation. Then, the fiber output signal is the waveform $a\tau_0 + u(t)$ where $u(t)$ is given by Eq. (3).

In order to create a heat channel, choose at the receiver any fixed time parameter α with the property that $\alpha > 1/\beta$. When the product $\alpha\beta$ is large, the dilation in (6) may be skipped in practice (for simplicity of exposition we imagine of having performed the dilation). Now, apply a variable density neutral filter of appropriate characteristic (an optical device, see [35]) to effect—in the spatial domain—a multiplication of the signal with the time window $q_\alpha(t) = \exp[-t^2/(2\alpha^2)]$ as in (6). Next, amplify the obtained signal by factor $1/a$. When an optical amplifier is employed, the resulting signal is $c_0 q_\alpha(t) + g(t) + n(t)$ where $c_0 = \tau_0(\cosh\delta)^{-1/2}$, $g = \boldsymbol{P}_\delta^{(\gamma)} f$, and $n(t)$ is a realization of white Gaussian noise (properly modeling the impairment of an optical signal by an optical amplifier; see [6]). After opto electric conversion (possibly adding new white Gaussian noise to the signal), remove the known signal component $c_0 q_\alpha(t)$. Finally, use as detection device a bank of matched filters (*cf.* [6]), with impulse responses $h_k(t)$ as in Section 2.2, $k = 0, 1, \ldots, K-1$, where K is a known number for any given average input energy S.

Thus, we have implemented a heat channel in an optical fiber communication system.

2.4 Degrees of Freedom of Filter Output Signals

Here, we give an explanation for the time-frequency product $\alpha\beta$ that will occur very frequently in the sequel.

The Wigner-Ville spectrum (WVS) of the response of filter $\boldsymbol{P}_\delta^{(\gamma)}$ on white Gaussian noise (see Appendix A for details) is the bivariate function

$$\boldsymbol{\Phi}(t, \omega) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{\cosh\delta} \exp\left(-\frac{t^2}{\alpha^2} - \frac{\omega^2}{\beta^2}\right). \tag{12}$$

In general, the WVS of a nonstationary stochastic process gives its density of (mean) energy in the time-frequency plane; see, *e.g.*, [27], [26]. Consequently, in our case, the energy of individual filter output signals would occupy an ellipse-shaped region in the time-frequency plane with unsharp boundary. Regarding the WVS (12) (after a normalization) as a bivariate Gaussian probability density function, we describe this region by an approximation known as ellipse of concentration (EoC) in probability theory [36]; the EoC has the property that

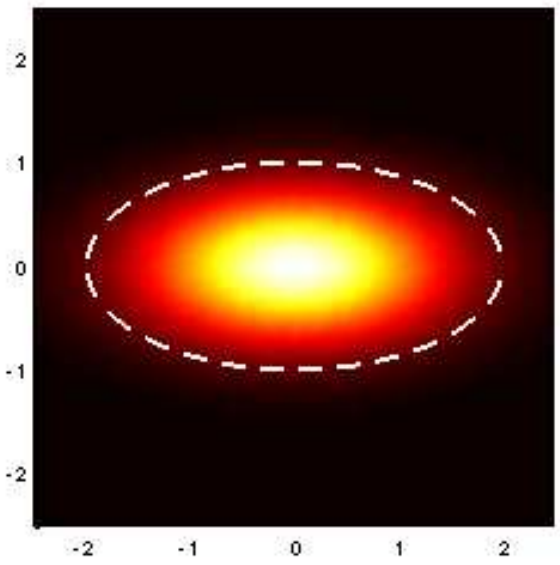

Figure 5: Energy density and ellipse of concentration (dashed line) for white Gaussian noise response of filter (4) in the limiting case $\alpha \to \sqrt{2}+$, $\beta \to (1/\sqrt{2})+$; graphical appearance typical also for admissible values of α, β.

the uniform distribution on it has the same first and second moments as the Gaussian distribution. In Fig. 5, the energy density in the time-frequency plane given by the WVS (12) is illustrated.

In our case, we obtain as EoC the region $\mathcal{A}_c = \{(t,\omega) \in \mathbb{R}^2; t^2/\alpha^2 + \omega^2/\beta^2 \leq 2\}$ with area $A_c = 2\pi\alpha\beta$ [14]. As reported in [12, p. 23], in physics a region in phase space (or time-frequency plane such as here) with area A corresponds to $A/(2\pi)$ "independent states" (when A is sufficiently large). In our case, we have $A_c/(2\pi) = \alpha\beta$. Accordingly, the time-frequency product $\alpha\beta$ would describe the "dimension" of the filter output space (when $\alpha\beta$ is sufficiently large), *i.e.*, the degrees of freedom (DoF) of filter output signals.

3 Channel Capacity in Terms of Channel Input, and Optimal Signaling

In this section, we derive a closed formula for the capacity of the heat channel in terms of average energy of the channel input signal along with a method of capacity achieving (optimal) signaling. A characterization of channel capacity by water-filling in the time-frequency plane is also given. The capacity results

will be stated in form of asymptotic equations.

Definition 2 *Any two functions A, $B : (1, \infty) \to \mathbb{R}$ are said to be asymptotically equal and we write $A \doteq B$, if*

$$\lim_{x\to\infty} \frac{A(x) - B(x)}{x} = 0,$$

or equivalently, $A(x) = B(x) + o(x)$ as $x \to \infty$.

In our context, x will always be the time-frequency product $\alpha\beta > 1$. Thus, $A \doteq B$ implies that $A(\alpha\beta)/(\alpha\beta) = B(\alpha\beta)/(\alpha\beta) + \epsilon$ where $\epsilon \to 0$ as $\alpha\beta \to \infty$.

3.1 Channel Capacity—First Formula

The function $y = w_0(x)$, $x \geq 0$, occurring in the next theorem is the inverse function of $y = (2x - 1)e^{2x} + 1$, $x \geq 0$ (see also Fig. 6).

Theorem 1 *Assume that the average energy S of the input signal depends on $\alpha\beta$ such that $S(\alpha\beta) = O(\alpha\beta)$ as $\alpha\beta \to \infty$. Then for the capacity (in bits per transmission) of the heat channel it holds*

$$C(S) \doteq \frac{\alpha\beta}{2} \left[w_0 \left(\frac{S}{(\alpha\beta/2)\theta^2} \right) \right]^2 \log_2 e. \tag{13}$$

Proof: The proof is accomplished by water-filling [10], [7, Theorem 7.5.1]. Let $\nu_k^2 = \theta^2 \rho^{-2k-1}$, $k = 0, 1, \ldots$, be the variance of noise in the kth subchannel. The positive number σ is defined by the condition

$$S = \sum_{k=0}^{\infty} (\sigma^2 - \nu_k^2)^+ = \sum_{k=0}^{K-1} (\sigma^2 - \nu_k^2), \tag{14}$$

where $x^+ \triangleq \max\{0, x\}$, $x \in \mathbb{R}$, and $K = \max\{k \in \mathbb{N}; \nu_{k-1}^2 < \sigma^2\}$ is the number of subchannels in the resulting finite-horizon version of the heat channel. With increasing time-frequency product $\alpha\beta$, $\delta = \delta(\alpha\beta)$ (now acting as increment) tends to 0 so that

$$\begin{aligned} S \cdot \delta &= \sum_{k=0}^{K-1} (\sigma^2 - \theta^2 e^{2k\delta} e^{\delta})\delta \\ &= \int_0^{\infty} (\sigma^2 - \theta^2 e^{2x})^+ \, dx + \epsilon, \end{aligned} \tag{15}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Observe that by the growth condition imposed on $S = S(\alpha\beta)$ and because of $\delta \sim \frac{1}{\alpha\beta}$, it holds $\limsup_{\alpha\beta\to\infty} S \cdot \delta < \infty$ so that transition to a Riemann integral is allowed; for later reference still note that the water level $\sigma^2 = \sigma^2(\alpha\beta)$ also remains bounded as $\alpha\beta \to \infty$. Evaluation of the integral yields

$$S \doteq \frac{\alpha\beta}{2}\theta^2 \left[\frac{\sigma^2}{\theta^2} \ln \left(\frac{\sigma^2}{\theta^2} \right) - \frac{\sigma^2}{\theta^2} + 1 \right]. \tag{16}$$

The maximum in Eq. (10) is achieved when the components X_k of the input vector $\boldsymbol{X}^K$ are independent $\sim \mathcal{N}(0, \sigma^2 - \nu_k^2)$ and the the capacity (in nats) becomes

$$C = C(S) = \sum_{k=0}^{K-1} \frac{1}{2} \ln \left(1 + \frac{\sigma^2 - \nu_k^2}{\nu_k^2} \right). \tag{17}$$

Since σ^2 eventually remains bounded, transition to a Riemann integral in the next equations is allowed and we get

$$\begin{aligned} C \cdot \delta &= \sum_{k=0}^{K-1} \frac{1}{2} \ln \left(\frac{\sigma^2}{\theta^2} e^{-2k\delta} e^{-\delta} \right) \delta \\ &= \int_0^{\infty} \left[\frac{1}{2} \ln \left(\frac{\sigma^2}{\theta^2} e^{-2x} \right) \right]^+ dx + \epsilon \\ &= \frac{1}{2} \left(\frac{1}{2} \ln \frac{\sigma^2}{\theta^2} \right)^2 + \epsilon, \end{aligned}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Thus it holds

$$C \doteq \frac{\alpha\beta}{2} \left(\frac{1}{2} \ln \frac{\sigma^2}{\theta^2} \right)^2. \tag{18}$$

Eq. (16) is equivalent to

$$\frac{\sigma^2}{e\theta^2} \ln \frac{\sigma^2}{e\theta^2} = e^{-1}(s-1) + \epsilon_1,$$

where $s = S/[(\alpha\beta/2)\theta^2]$ and $\epsilon_1 \to 0$ as $\alpha\beta \to \infty$. By means of the Lambert W function [37] (actually its principal branch W_0, see Fig. 6) which is the uniquely determined analytic function satisfying $W(x)\exp[W(x)] = x$ for all $x \in [-e^{-1}, \infty)$ and $W(0) = 0$, we get after a computation

$$\frac{1}{2} \left(\frac{1}{2} \ln \frac{\sigma^2}{\theta^2} \right)^2 = \frac{1}{2} \left[w_0(s + \epsilon_1') \right]^2,$$

where we have set $w_0(x) = \frac{1}{2}[1 + W((x-1)/e)]$, $x \geq 0$, and $\epsilon_1' = e\epsilon_1$.

Because of Eq. (18), this gives rise to

$$\frac{C}{\alpha\beta} = \frac{1}{2} \left[w_0(s + \epsilon_1') \right]^2 + \epsilon_2,$$

where $\epsilon_2 \to 0$ as $\alpha\beta \to \infty$. Unlike $w_0(x)$, which has a vertical tangent at $x = 0$, the function $(w_0(x))^2$ has a continuous and bounded derivative in every closed interval $[0, s_0] \subset [0, \infty)$. Consequently, by the mean value theorem we obtain $C/(\alpha\beta) = \frac{1}{2}\left[w_0(s)\right]^2 + \epsilon_1'' + \epsilon_2$ where $\epsilon_1'' + \epsilon_2 \to 0$ as $\alpha\beta \to \infty$. After transition from nats to bits ($1\,\text{nat} = \log_2 e$ bit), Eq. (13) is obtained. □

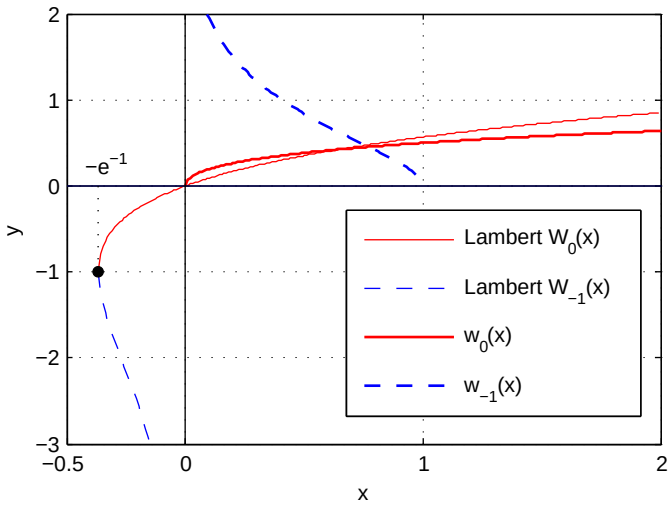

Figure 6: Lambert W function: branches $W_0(x)$, $W_{-1}(x)$ and related functions $w_0(x)$, $w_{-1}(x)$.

Remark 1 *In [38] the increment $1/K$ for a similar transition to an integral as in (15) is used (the parameter K in [38] corresponds to the number K of subchannels in our paper). This would not be an appropriate choice in the above proof, since in our case [see also Eq. (24), below] K depends on input energy (or SNR).*

Remark 2 *In the above proof, the heat channel was reduced to a finite-dimensional vector Gaussian channel. For such channels, a coding theorem is available [7, Theorem 7.5.2], [10], i.e., the capacity $C(S)$ in Eq. (17) can be achieved by a sequence of codes. Therefore, the capacity given by Eq. (13) has also an operational meaning.*

We discuss the case $0 < S = S(\alpha\beta) \propto \alpha\beta$ in more detail. First, when $\beta > 0$ is held constant, then $S = \alpha\beta S_1 = \alpha P$, $P = \beta S_1$. Assume that one transmission takes time α (in seconds). Forming the limit $\lim_{\alpha\to\infty} C(S)/\alpha = \bar{C}$ turns Eq. (13) into the true equation

$$\bar{C} = \frac{\beta}{2}\left[w_0\left(\frac{P}{(\beta/2)\theta^2}\right)\right]^2 \log_2 e \ \text{(bit/s)}. \tag{19}$$

Next, let $\beta \to \infty$. Since $w_0(0) = 0$ and $(w_0(x))^2$ is differentiable at $x = 0$ with derivative $1/2$, it follows that

$$\bar{C} \to \frac{P}{2\theta^2} \log_2 e \ \text{(bit/s)}. \tag{20}$$

Finally, we compare (19) with the capacity of the bandlimited Gaussian channel of bandwidth W and one-sided noise PSD N_0 given by Shannon's classic

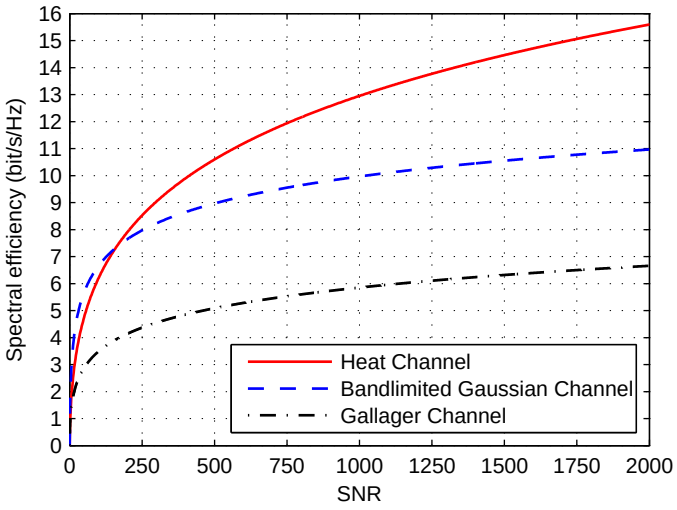

Figure 7: Spectral efficiencies of heat channel, bandlimited Gaussian channel, and Gallager channel as a function of SNR. For small SNR, the spectral efficiency of the heat channel is smaller than that of the Gallager channel which cannot be seen in the present resolution.

formula [9]

$$C = W \log_2 \left(1 + \frac{P}{WN_0}\right) \text{ (bit/s)}. \tag{21}$$

Rewrite the latter equation as $C(\text{SNR}) = W \log_2(1 + \text{SNR})$ where $\text{SNR} = P/(WN_0)$. In case of the heat channel, it is consistent to set $\text{SNR} = P/(WN_0)$ where $W = \beta/2$ (*cf.* Section 1), $N_0 = 2\theta^2$ is the *one*-sided noise PSD (*cf.* Section 2.2), and to rewrite Eq. (19) as $C(\text{SNR}) = W\ [w_0\,(2 \cdot \text{SNR})]^2 \log_2 e$. In Fig. 7, the corresponding spectral efficiencies $C(\text{SNR})/W$ are plotted as a function of SNR. As underpinned by Fig. 8, we observe that the spectral efficiency of the heat channel eventually becomes larger than that of the bandlimited Gaussian channel. On the other hand, the capacity limit in (20) is exactly the same as for a Gaussian channel with infinite bandwidth, average input power P and one-sided noise PSD $N_0 = 2\theta^2$ (*cf.*, *e.g.*, [10, (9.63)]).

3.2 Optimal Signaling for the Continuous-Time Channel

We return to our original, by now refined continuous-time model of the heat channel consisting of the LTV filter (4) as its first component, followed by AWGN, and a final measurement step. For a fixed average input energy $S > 0$ and noise variance $\theta^2 > 0$, the capacity achieving (optimal) input signals are now waveforms

$$f(t) = \sum_{k=0}^{K-1} x_k\,(D_\gamma \psi_k)(t),\ t \in \mathbb{R}, \tag{22}$$

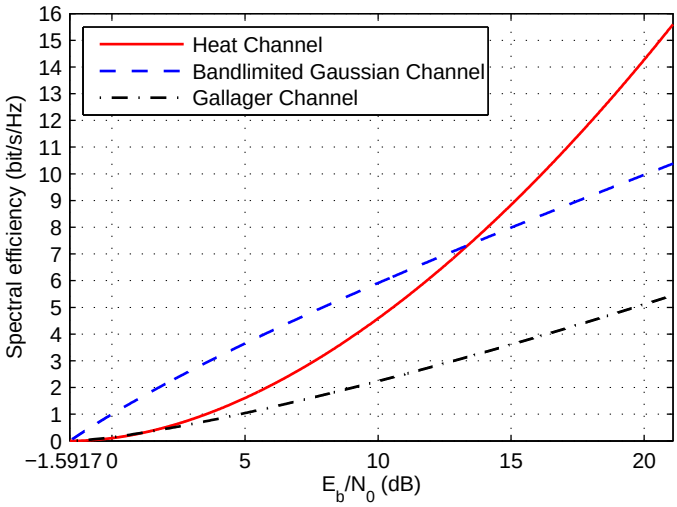

Figure 8: Spectral efficiencies of heat channel, bandlimited Gaussian channel, and Gallager channel plotted against $10 \log_{10} E_\mathrm{b}/N_0$; E_b is average input energy per bit, N_0 one-sided noise PSD of AWGN. In the vicinity of the origin, the spectral efficiency of the heat channel is smaller than that of the Gallager channel which is hard to be seen in the present resolution.

where the coefficients x_k are realizations of independent Gaussian random variables $X_k \sim \mathcal{N}(0, \sigma^2 - \nu_k^2)$, $k = 0, \ldots, K-1$, as given in the proof of Theorem 1. The corresponding ("optimal") filter output signal is

$$g(t) = \sum_{k=0}^{K-1} y_k \, (D_\gamma \psi_k)(t), \; t \in \mathbb{R}, \tag{23}$$

where the coefficients $y_k = \rho^{k+\frac{1}{2}} x_k$ are realizations of independent Gaussian random variables $Y_k \sim \mathcal{N}(0, \sigma_k^2 - \theta^2)$, $\sigma_k^2 = \sigma^2 \rho^{2k+1}$, $k = 0, \ldots, K-1$.

The following numerical example may serve as illustration. In Fig. 9, a pair of optimal filter input/output signals is displayed. Since transmission through an optical fiber by intensity modulation is supposed, the input signal $f(t)$ is modified as described in Section 2.3. By means of Eq. (13), the capacity of the concrete heat channel is found to be approximately 29.47 bits per transmission (in the given precision, this figure agrees with the capacity found by numerical computation). Observe that the input signal $f(t)$ in Fig. 9 is practically of finite duration and confined to the time interval $[-\pi\alpha, \pi\alpha]$ (or its spatial equivalent) as is the filter output signal $g(t)$. Now assume that a signal such as $f(t)$ is generated and sent every time α, the average input power then being $P = S/\alpha$. When the ratio x/t for conversion from the temporal (electrical) domain to the spatial (optical) domain is 2π m/s, a signal of effective duration of at most 1 second (as in the example in Fig. 9) would result in an optical waveform of

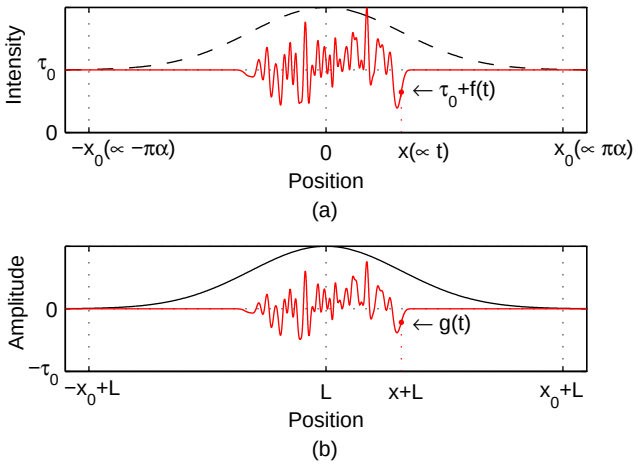

Figure 9: Optimal waveforms for the heat channel with parameters $\alpha = 1$ s, $\beta = 100$ Hz, $\theta^2 = 0.01$ W/Hz, $S = 1.0$ Ws. (a) Physical input signal $\tau_0 + f(t)$, $\tau_0 = 2$ [no unit], to an optical fiber of length L; intensity modulation assumed. (b) Corresponding filter output signal $g(t)$. Gaussian time window enlarged. Signals displayed in spatial domain; ratio x/t order of 1 m/s, say.

length of at most 6.28 meters. Then, given a speed of light of $200,000$ km/s in the fiber, an interference of waveforms in the fiber would, of course, be of no importance at all. As a consequence, a rate of 29.47 bit/s would be attained; the same value is found by Eq. (19). In the present example, $K = 64$ subchannels are needed (as numerically determined).

Now, let $\alpha\beta \to \infty$. Under the assumption of Theorem 1, the number K of active subchannels is found to be

$$K \doteq \frac{\alpha\beta}{2} \ln \frac{\sigma^2}{\theta^2} \,. \tag{24}$$

For the sake of simplicity, assume further as above that $0 < S(\alpha\beta) \propto \alpha\beta$. Then, by (15), we infer that σ^2 approaches a finite water level $\bar{\sigma}^2 > \theta^2$ as $\alpha\beta \to \infty$. As a consequence, $K \to \infty$ and $\sigma_k^2 - \theta^2 \to \bar{\sigma}_k^2 - \theta^2$ where we have set $\bar{\sigma}_k^2 = \bar{\sigma}^2 \rho^{2k+1}$. When $\bar{\sigma}^2$ is large compared to θ^2, the bias θ^2 in $\bar{\sigma}_k^2 - \theta^2$ may be neglected so that the optimal filter output signal (23) comes close to a white Gaussian noise response; *cf.* (32) (below) and Appendix A. Then, the signal model of Section 2.4 is almost met and we may take (and shall do so in the sequel) the time-frequency product $\alpha\beta$ as DoF of optimal filter output signals.

3.3 Channel Capacity by Water-Filling in the Time-Frequency Plane, and a Classic Result of Gallager.

By means of a Szegö theorem (*viz.* Theorem 6 in Appendix B), the above water-filling solution carries over to the time-frequency plane. The classic water-filling solution for the capacity of a power-constraint additive Gaussian noise channel goes back to Shannon [39] and has been stated and proved by Gallager [7] in full generality.

3.3.1 Capacity of the Heat Channel by Water-Filling in Time and Frequency

Referring to Gallager's result [7, Theorem 8.5.1] in the form given in [10, (9.97)], we define

$$N(t,\omega) = \frac{\theta^2}{2\pi} \cdot (\cosh \delta) \exp\left(\frac{t^2}{\alpha^2} + \frac{\omega^2}{\beta^2}\right).$$

Now we are in a position to state

Theorem 2 *Under the same assumption on the average input energy S as in Theorem 1, the capacity (in bits per transmission) of the heat channel is given by*

$$C \doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} \frac{1}{2} \log_2 \left(1 + \frac{(\nu - N(t,\omega))^+}{N(t,\omega)}\right) dt\, d\omega, \tag{25}$$

where ν is chosen so that

$$S \doteq \iint_{\mathbb{R}^2} (\nu - N(t,\omega))^+ \, dt\, d\omega. \tag{26}$$

Proof: See Appendix B. □

Note that the bivariate function $N(t,\omega)$ is proportional to the reciprocal WVS $\mathbf{\Phi}(t,\omega)$ in (12). Since $N(t,\omega)$ has the form of a "cup," Theorem 2 is a water-filling theorem in a very real sense.

3.3.2 Comparison with Gallager's Result

When $\alpha \to \infty$, the continuous-time heat channel [*cf.* Eq. (6) and Fig. 2(a)] appears to tend towards an LTI channel according to Gallager's model[3] with LTI filter with impulse response (2). It is therefore worthwhile to compare Theorem 1 and 2 with Gallager's classic result [7, Theorem 8.5.1] when the latter is applied to that particular LTI channel. According to Gallager's theorem, the capacity C (now a rate measured in bits per second) at input power P is given

[3] In the statement of [7, Theorem 8.5.1] the crucial assumption "$T \to \infty$" is missing. As a consequence, in [7, Figure 8.5.1] the restriction of the input $x(t)$ to a bounded time interval $(-T/2, T/2)$ would drop out. By this reason, in Fig. 2(a) any time constraint on the input is omitted.

parametrically by

$$\begin{aligned} C &= \int_{-\infty}^{\infty} \frac{1}{2} \left[\log_2 \frac{|H_1(f)|^2 B}{N_0/2} \right]^+ df \\ P &= \int_{-\infty}^{\infty} \left[B - \frac{N_0/2}{|H_1(f)|^2} \right]^+ df, \end{aligned}$$

where $H_1(f) = \int_{-\infty}^{\infty} e^{-i2\pi ft} h_1(t)\, dt$ is the frequency response of the filter and $N_0/2$ is the two-sided noise PSD of the AWGN. For the function $h_1(t)$ at hand and AWGN with $N_0/2 = \theta^2$ (*cf.* Section 2.2), we obtain after a calculation

$$C = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{2} \log_2 \left(1 + \frac{(\nu_1 - N_1(\omega))^+}{N_1(\omega)} \right) d\omega \tag{27}$$

$$P = \int_{-\infty}^{\infty} (\nu_1 - N_1(\omega))^+ \, d\omega, \tag{28}$$

where $\nu_1 = \frac{B}{2\pi}$ is the new parameter, $\omega = 2\pi f$ is the integration variable, and

$$N_1(\omega) = \frac{\theta^2}{2\pi} \cdot \exp\left(\frac{\omega^2}{\beta^2} \right). \tag{29}$$

We observe formal analogy between the preceding water-filling formulas (27), (28) and (25), (26) in Theorem 2. Moreover, it holds that $N(t, \omega) \to N_1(\omega)$ as $\alpha \to \infty$ for any t, ω held constant.

Now, from (27) and (28) we readily obtain

$$\frac{C}{\beta/2} = \frac{2}{3\pi} (\ln r)^{\frac{3}{2}} \log_2 e \quad \text{bit/s/Hz} \tag{30}$$

$$\frac{P}{\beta\theta^2} = \frac{1}{\pi} \int_0^{\infty} \left(r - e^{x^2} \right)^+ dx, \tag{31}$$

where $r = B/\theta^2$ is again a new parameter. With the setting $\mathrm{SNR} = P/(WN_0)$ where $W = \beta/2$, $N_0 = 2\theta^2$ (same as in Section 3.1 in case of the heat channel) we thus obtain a parametric representation of a curve in the C/W, SNR plane. In Fig. 7, the spectral efficiency C/W is plotted as a function of SNR. It came as a big surprise to us that the spectral efficiency of the above Gallager channel is inferior to that of the heat channel in the limiting case $\alpha \to \infty$ (as inferred from Theorem 1)—at least when SNR is not too small. A similar behaviour is to be observed in Fig. 8.

We still remark that it seems elusive to find a closed-form expression from Eqs. (30), (31) for C/W as a function of SNR; suffice it to say that, to the best of our knowledge, the integral in (31) cannot be expressed in terms of elementary functions [whereas it is easy to evaluate the double integral in (26)].

4 Rate Distortion Function of a Connected Gaussian Process

For a well-rounded treatment of the capacity problem for the heat channel it is expedient to investigate a dual problem, which is a topic of rate distortion theory. To this end, consider the nonstationary zero-mean Gaussian process defined by the Karhunen-Loève expansion

$$X(t) = \sum_{k=0}^{\infty} X_k \, (D_\gamma \psi_k)(t), \; t \in \mathbb{R}, \tag{32}$$

where the coefficients X_k, $k = 0, 1, \ldots$, are independent Gaussian random variables $\sim \mathcal{N}(0, \sigma_k^2)$ of variance $\sigma_k^2 = \sigma^2 \rho^{2k+1}$, $\sigma > 0$; it is the response of filter $\boldsymbol{P}_\delta^{(\gamma)}$ on white Gaussian noise [*cf.* (52) in Appendix A]. In Fig. 4, the area beneath the curve $y = \sigma^2 e^{-2x}$ corresponds to the average energy

$$E = \sum_{k=0}^{\infty} \sigma_k^2 = \frac{\sigma^2 \rho}{1 - \rho^2} \doteq \frac{\alpha\beta}{2} \sigma^2 \tag{33}$$

of the Gaussian process (32). The parameter θ^2 in Fig. 4 will now have the interpretation of a "(ground-)water table;" in the concept of *test channel* [10, p. 311], it would have the meaning of noise variance again.

In this section, information will be measured in nats.

4.1 Rate Distortion Function in Closed Form

Substitute the continuous-time Gaussian process $\{X(t), \, t \in \mathbb{R}\}$ in (32) by the sequence of coefficient random variables $\boldsymbol{X} = X_0, X_1, \ldots$. For an estimate $\hat{\boldsymbol{X}} = \hat{X}_0, \hat{X}_1, \ldots$ of $\boldsymbol{X}$ we take the mean-square error $D = \mathbb{E}\{\sum_{k=0}^{\infty} (X_k - \hat{X}_k)^2\}$ as distortion measure.

The function $y = w_{-1}(x)$, $0 < x \leq 1$, occurring in the next theorem is the inverse function of $y = (2x+1)e^{-2x}$, $x \geq 0$ (see also Fig. 6). The Landau symbol $\Omega(\cdot)$ is defined for any two functions as in Definition 2 as follows: $A(x) = \Omega(B(x))$ as $x \to \infty$ if $B(x) > 0$ and $\liminf_{x \to \infty} A(x)/B(x) > 0$.

Theorem 3 *Assume that the foregoing average distortion D depends on $\alpha\beta$ such that $D(\alpha\beta) = \Omega(\alpha\beta)$ as $\alpha\beta \to \infty$. Then the rate distortion function of the Gaussian process (32) satisfies*

$$R(D) \doteq \frac{\alpha\beta}{2} \left[w_{-1} \left(\frac{D}{(\alpha\beta/2)\sigma^2} \right) \right]^2 \tag{34}$$

if $D \leq (\alpha\beta/2)\sigma^2$, and $R(D) \doteq 0$ otherwise.

Proof: Let E be the average energy (33) of the Gaussian process (32).

First, assume $D \leq E$. The reverse water-filling argument for a finite number of independent Gaussian sources [10] carries over to our case without changes

resulting in a finite collection of Gaussian sources $X_0, \ldots, X_{K-1}$ where $K = \max\{k \in \mathbb{N}; \sigma_{k-1}^2 > \theta^2\}$ and the water level $\theta^2 > 0$ is defined by the condition

$$D = \sum_{k=0}^{\infty} \min\{\theta^2, \sigma_k^2\}, \tag{35}$$

(*cf.* Fig. 4, where $D = E_{\text{dist1}} + E_{\text{dist2}}$). Consequently,

$$\begin{aligned} D \cdot \delta &= \sum_{k=0}^{\infty} \min\{\theta^2, \sigma^2 e^{-2k\delta} e^{-\delta}\}\delta \\ &= \int_0^{\infty} \min\{\theta^2, \sigma^2 e^{-2x}\}\, dx + \epsilon, \end{aligned} \tag{36}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Observe that by the growth condition imposed on D and since $\delta \sim \frac{1}{\alpha\beta}$, the water level θ^2 eventually remains above a positive lower bound as $\alpha\beta \to \infty$. Evaluation of the integral yields

$$D \doteq \frac{\alpha\beta}{2}\theta^2 \left[\ln\left(\frac{\sigma^2}{\theta^2}\right) + 1\right]. \tag{37}$$

The rate distortion function is parametrically given by [10]

$$R = \sum_{k=0}^{K-1} \frac{1}{2} \ln \frac{\sigma_k^2}{\theta^2}. \tag{38}$$

The RHSs of Eqs. (38) and (17) agree. Since $\frac{1}{\theta^2}$ is eventually finitely upper bounded, transition to a Riemann integral is allowed and we obtain exactly as in the proof of (18) that

$$R \doteq \frac{\alpha\beta}{2} \left(\frac{1}{2} \ln \frac{\sigma^2}{\theta^2}\right)^2. \tag{39}$$

Eq. (37) is equivalent to

$$\frac{\theta^2}{e\sigma^2} \ln \frac{\theta^2}{e\sigma^2} = -e^{-1}d + \epsilon_1,$$

where $d = D/[(\alpha\beta/2)\sigma^2]$ and $\epsilon_1 \to 0$ as $\alpha\beta \to \infty$. By means of the branch W_{-1} of the Lambert W function [37], which is the uniquely determined analytic function satisfying $W_{-1}(x)\exp[W_{-1}(x)] = x$ for all $x \in [-e^{-1}, 0)$ and $W_{-1}(-e^{-1}) = -1$, $W_{-1}(x) \to -\infty$ as $x \to 0-$ (see Fig. 6), we get after a computation

$$\frac{1}{2}\left(\frac{1}{2} \ln \frac{\sigma^2}{\theta^2}\right)^2 = \frac{1}{2}\left[w_{-1}(d + \epsilon_1')\right]^2,$$

where we have set $w_{-1}(x) = \frac{1}{2}[-1 - W_{-1}(-x/e)]$, $0 < x \leq 1$, and $\epsilon_1' = -e\epsilon_1$. Because of Eq. (39), this gives rise to

$$\frac{R}{\alpha\beta} = \frac{1}{2}\left[w_{-1}(d + \epsilon_1')\right]^2 + \epsilon_2,$$

where $\epsilon_2 \to 0$ as $\alpha\beta \to \infty$. Unlike $w_{-1}(x)$, which has a vertical tangent at $x = 1$, the function $(w_{-1}(x))^2$ has a continuous and bounded derivative in any closed interval $[d_0, 1] \subset (0, 1]$. Consequently, by the mean value theorem $R/(\alpha\beta) = \frac{1}{2}\left[w_{-1}(d)\right]^2 + \epsilon_1'' + \epsilon_2$ where $\epsilon_1'' + \epsilon_2 \to 0$ as $\alpha\beta \to \infty$ uniformly for all $d \in [d_0, 1]$. This proves the first part of the theorem.

Now, suppose $D > E$. Since $\mathbb{E}\{\sum_{k=0}^{\infty}(X_k - 0)^2\} = E < D$, the constant sequence $\hat{\boldsymbol{X}} = 0, 0, \ldots$ is a sufficient estimate. Since there is no uncertainty about the members of that deterministic sequence, no information needs to be supplied; thus, $R = 0$. This proves the second part of the theorem. □

Observe that although the functions $w_{-1}(x)$ in Eq. (34) and $w_0(x)$ in Eq. (13) seem to be completely different, they are linked by the two branches of the Lambert W function (*cf.* Fig. 6).

4.2 Reverse Water-Filling in the Time-Frequency Plane

Before continuing with our main theme, we present a parametric representation of the rate distortion function occurring in Theorem 3 since the means for its proof are now available (*cf.* Theorem 6 in Appendix B). This representation may be viewed as an extension to the time-frequency plane of the classic method of reverse water-filling (*cf.*, *e.g.*, [28], [10]) due to Kolmogorov [40]. It turns out that the part of the PSD in [28, Theorem 4.5.4] is now taken by the WVS $\boldsymbol{\Phi}(t, \omega)$ in (12). We obtain

Theorem 4 *The rate distortion function $R(D)$ of the nonstationary Gaussian process (32) has in the interval $0 < D \leq (\alpha\beta/2)\sigma^2$ the parametric representation*

$$\begin{aligned} D_\lambda &\doteq \iint_{\mathbb{R}^2} \min\left\{\lambda, \boldsymbol{\Phi}(t, \omega)\right\}\, dt\, d\omega \\ R_\lambda &\doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} \max\left\{0, \frac{1}{2}\ln\left(\frac{\boldsymbol{\Phi}(t, \omega)}{\lambda}\right)\right\}\, dt\, d\omega. \end{aligned}$$

Proof: See Appendix B. □

Note that

$$\iint_{\mathbb{R}^2} \boldsymbol{\Phi}(t, \omega)\, dt\, d\omega = \frac{\alpha\beta}{2}\frac{\sigma^2}{\cosh\delta} \doteq \frac{\alpha\beta}{2}\sigma^2$$

is the average energy (33) of the nonstationary Gaussian process (32)—as it should be. We observe that the representation in Theorem 4 is in perfect analogy to the parametric representation [28, (4.5.51), (4.5.52)] of the $R(D)$ function of a continuous-time stationary Gaussian source.

5 Channel Capacity in Terms of Channel Output, and Relation to Estimation Theory

Now, we adopt the perspective of the receiver. This will result in a second closed-form capacity formula for the heat channel, now in terms of elementary functions

and akin to the classic Shannon formula (21) for the capacity of a bandlimited Gaussian channel. Moreover, we shall find a parallel to the celebrated I-MMSE relationship [see Eq. (44) below] due to Guo *et al.* [29]. Implications for the capacity of multiple-input multiple-output (MIMO) systems will be indicated. In the present section, it is convenient to use natural logarithms; therefore, unless otherwise stated, information is measured in nats.

5.1 Channel Capacity—Second Formula

Recall the representation (23) of an optimal filter output signal g. In the next theorem, the capacity of the heat channel will be expressed in terms of average energy $\hat{E}_{\text{out}}$ of the measured filter output signal

$$\hat{g}(t) = \sum_{k=0}^{K-1} \hat{y}_k \, (D_\gamma \psi_k)(t), \; t \in \mathbb{R},$$

where $\hat{y}_k = y_k + n_k$, $y_k = \rho^{k+\frac{1}{2}} x_k$, and x_k is as in (22); the measurement errors n_k are realizations of i.i.d. Gaussian random variables $N_k \sim \mathcal{N}(0, \theta^2)$ (as in Section 2.2). Notice that here we assumed the model of an inaccurate measurement of the coefficients y_k of the noiseless filter output signal (23); we shall maintain this model of measurement in Section 5 throughout.

Theorem 5 *Assume that the average energy $\hat{E}_{\text{out}}$ of the measured filter output signal depends on $\alpha\beta$ such that $\hat{E}_{\text{out}}(\alpha\beta) = O(\alpha\beta)$ as $\alpha\beta \to \infty$. Then for the capacity (in bits per transmission) of the heat channel it holds*

$$C \doteq \frac{\alpha\beta}{2} \left[\ln \sqrt{1 + \frac{\hat{E}_{\text{out}}}{(\alpha\beta/2)\theta^2}} \right]^2 \log_2 e. \tag{40}$$

Proof: Rewrite Eq. (18) as

$$C \doteq \frac{\alpha\beta}{2} \left[\ln \sqrt{1 + \frac{(\alpha\beta/2)(\sigma^2 - \theta^2)}{(\alpha\beta/2)\theta^2}} \right]^2 \text{(nat)} \tag{41}$$

and set $\sigma_k^2 = \sigma^2 \rho^{2k+1}$, $k = 0, 1, \ldots$. In case of optimal signaling, the average energy of the measured filter output signal is $\hat{E}_{\text{out}} = E_{\text{out}} + E_{\text{err}}$ (*cf.* Fig. 4) where $E_{\text{out}} = \sum_{k=0}^{K-1} (\sigma_k^2 - \theta^2)$ is the average energy of the (noiseless) filter output signal, $E_{\text{err}} = \sum_{k=0}^{K-1} \theta^2$ is the average energy of the measurement error, and K is given by $K = \max\{k \in \mathbb{N}; \sigma_{k-1}^2 > \theta^2\}$ (coinciding with the number K of active subchannels in the proof of Theorem 1).

Since $\hat{E}_{\text{out}} = \sum_{k=0}^{K-1} \sigma_k^2$ and $\delta \sim \frac{1}{\alpha\beta}$, we get

$$\begin{aligned} \hat{E}_{\text{out}} \cdot \delta &= e^{-\delta} \sum_{k=0}^{K-1} \sigma^2 e^{-2k\delta} \delta \\ &= \int_0^{\frac{1}{2}\ln\frac{\sigma^2}{\theta^2}} \sigma^2 e^{-2x}\, dx + \epsilon \\ &= \frac{1}{2}(\sigma^2 - \theta^2) + \epsilon, \end{aligned}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Observe that by the growth condition imposed on $\hat{E}_{\text{out}}$, σ^2 remains bounded as $\alpha\beta \to \infty$ justifying in return the transition to a Riemann integral. Hence,

$$\hat{E}_{\text{out}} \doteq \frac{\alpha\beta}{2}(\sigma^2 - \theta^2). \tag{42}$$

Now, Eq. (40) follows from Eq. (41) and Eq. (42) (after final transition from nats to bits). □

Note that for the determination of channel capacity by formula (40), the receiver does not need to know the number K of active subchannels beforehand, since, at least in principle, it could easily be estimated as accurately as desired from successive optimal channel uses at constant average input energy.

In the rest of this section we shall always suppose that K is asymptotically given by Eq. (24), tacitly assuming that the assumptions of Theorem 1 (or Theorem 5) are fulfilled.

5.2 Relation to Estimation Theory

The setting of the previous subsection gives rise to a vector Gaussian channel

$$\boldsymbol{Y}^K = \boldsymbol{H}^K \boldsymbol{X}^K + \boldsymbol{N}^K, \tag{43}$$

where the matrix $\boldsymbol{H}^K$ is the $K \times K$ diagonal matrix with entries $h_{kk} = \rho^{k+\frac{1}{2}}$, $k = 0, \ldots, K-1$, $\boldsymbol{X}^K$ is a random input vector, $\boldsymbol{N}^K$ a noise vector with random components N_k as above, and $\boldsymbol{Y}^K$ the resulting output vector. In the language of vectors, in Section 5.1 the vector $\hat{\boldsymbol{y}}^K = (\hat{y}_0, \ldots, \hat{y}_{K-1})^T$ was obtained by measurement of the (noiseless) vector $\boldsymbol{a}^K = \boldsymbol{H}^K \boldsymbol{x}^K$ where $\boldsymbol{x}^K = (x_0, \ldots, x_{K-1})^T$ was the input. Conversely, we could think of $\hat{\boldsymbol{y}}^K$ as a perturbed version of $\boldsymbol{a}^K$ and aim to estimate $\boldsymbol{a}^K$ from $\hat{\boldsymbol{y}}^K$. Linear least squares estimation would result in $\hat{\boldsymbol{y}}^K$ itself with squared error $\|\boldsymbol{a}^K - \hat{\boldsymbol{y}}^K\|^2 = \sum_{k=0}^{K-1} (a_k - \hat{y}_k)^2$. Taking the expectation, the linear least squares error (LLSE) becomes

$$\mathbb{E}\{\|\boldsymbol{H}^K \boldsymbol{X}^K - \boldsymbol{Y}^K\|^2\} = \mathbb{E}\{\|\boldsymbol{N}^K\|^2\} = K\theta^2,$$

which, of course, agrees with the average energy of the measurement error $E_{\text{err}} = \sum_{k=0}^{K-1} \theta^2$ (*cf.* Fig. 4). So, we may interprete measurement as an estimation problem.

5.2.1 C-LLSE formula

The central result of the paper [29] is an identity connecting mutual information with the minimum mean-square error (MMSE) of estimation theory. In case of a Gaussian input, this identity reads

$$\frac{d}{d\,\mathrm{snr}} I(\boldsymbol{X}; \sqrt{\mathrm{snr}}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}) = \frac{1}{2}\mathrm{mmse}(\mathrm{snr}), \tag{44}$$

where $\boldsymbol{N}$ is a noise vector with independent standard Gaussian components, independent of the random vector $\boldsymbol{X}$, $\mathbb{E}\{\|\boldsymbol{X}\|^2\} < \infty$, and $\boldsymbol{H}$ is a deterministic matrix of appropriate dimension. It is interesting to compare Eq. (44) with the capacity calculations in our paper. We shall denote the inverse and transpose matrix of $\boldsymbol{H}^K$ by $\boldsymbol{H}^{-K}$ and $\boldsymbol{H}^{KT}$, respectively. Since mutual information is invariant with respect to invertible linear transformations, we infer for the mutual information $I(\boldsymbol{X}^K; \boldsymbol{Y}^K) = I(\boldsymbol{X}^K; \boldsymbol{X}^K + \boldsymbol{H}^{-K}\boldsymbol{N}^K)$ occurring in Eq. (10) that

$$I(\boldsymbol{X}^K; \boldsymbol{Y}^K) = I\left(\boldsymbol{X}'^K; \frac{\sigma}{\theta}\boldsymbol{H}^K\boldsymbol{X}'^K + \boldsymbol{N}'^K\right), \tag{45}$$

where the noise vector $\boldsymbol{N}'^K = \theta^{-1}\boldsymbol{N}^K$ has independent standard Gaussian components, independent of the random vector $\boldsymbol{X}'^K = \sigma^{-1}\boldsymbol{X}^K$. If we take $\boldsymbol{X}'^K = (X'_0, \ldots, X'_{K-1})^T$ with independent components $X'_k \sim \mathcal{N}(0, 1-(\theta^2/\sigma^2)\rho^{-2k-1})$, where σ is determined by Eq. (14) or, equivalently, $\sum_{k=0}^{K-1}[1-(\theta^2/\sigma^2)\rho^{-2k-1}] = S/\sigma^2$, then the left-hand side of Eq. (45) achieves capacity $C(S)$ of the heat channel. Since $C(S)$ depends only on the signal-to-noise ratio $\mathrm{snr} = \sigma^2/\theta^2 \in [1, \infty)$,[4] we may write (with slight abuse of notation)

$$C(\mathrm{snr}) = I(\boldsymbol{X}'^K; \sqrt{\mathrm{snr}}\boldsymbol{H}^K\boldsymbol{X}'^K + \boldsymbol{N}'^K), \tag{46}$$

which is reminiscent of the mutual information in (44).

Now, several problems arise when trying to take the derivate with respect to snr: 1) The probability distribution of the input vector $\boldsymbol{X}'^K$ depends on snr, 2) The function $C(\mathrm{snr})$ is not differentiable at snrs where a new subchannel is added (*cf.* Fig. 10). To overcome both problems, we substitute $C(\mathrm{snr})$ by its smooth approximation $C_0(\mathrm{snr}) = \frac{\alpha\beta}{2}\left(\ln\sqrt{\mathrm{snr}}\right)^2$ as given by Eq. (18) of Section 3.1. Then we may state what could be called a C-LLSE formula in the following theorem.

Proposition 1 *Assume* $\sigma^2 = 1$ *and* $0 < \theta^2 \leq 1$. *Then it holds that*

$$\frac{d}{d\,\mathrm{snr}} C_0(\mathrm{snr}) \doteq \frac{1}{2}\mathrm{llse}(\mathrm{snr}), \tag{47}$$

where $\mathrm{llse}(\mathrm{snr})$ *is the LLSE at* $\mathrm{snr} = \sigma^2/\theta^2 = 1/\theta^2$.

[4] Since only the portion $\sigma^2 - \theta^2$ contributes to the signal, σ^2/θ^2 is rather a *signal plus noise* -to-noise ratio; we stick to the notation "snr" to conform with [29].

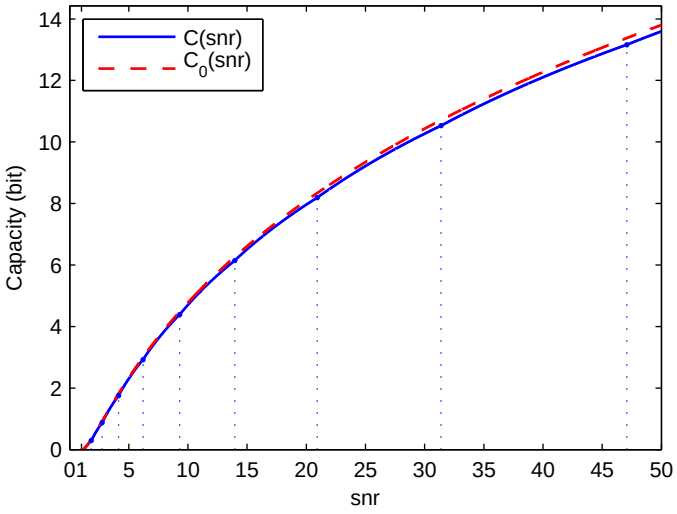

Figure 10: Capacity $C(\mathrm{snr})$ of heat channel, and its approximation $C_0(\mathrm{snr})$ in case $\alpha\beta = 5$ (for larger values of $\alpha\beta$, the two curves quickly become indistinguishable in the given range of snr). Dotted lines depict snrs where differentiability of $C(\mathrm{snr})$ breaks down.

Proof: First, let σ, θ be arbitrary numbers with $0 < \theta \leq \sigma < \infty$. Taking the derivative of $C_0(\mathrm{snr})$ with respect to $\mathrm{snr} = \sigma^2/\theta^2$ we get, observing Eq. (24),

$$\frac{d}{d\,\mathrm{snr}} C_0(\mathrm{snr}) = \frac{1}{2}\frac{\theta^2}{\sigma^2} \cdot \frac{\alpha\beta}{2} \ln \frac{\sigma^2}{\theta^2} \doteq \frac{1}{2}\frac{K\theta^2}{\sigma^2} = \frac{1}{2}\frac{E_{\mathrm{err}}}{\sigma^2}.$$

After rescaling $1, \theta/\sigma \leftarrow \sigma, \theta$ the snr is retained and we obtain (47) where $\mathrm{llse}(\mathrm{snr}) = K(\theta/\sigma)^2 \leftarrow E_{\mathrm{err}}$ is the new LLSE at the given snr. □

After all, Eq. (47) is similar to Eq. (44) again! We remark that the LLSE in Proposition 1 is given by

$$\mathrm{llse}(\mathrm{snr}) \doteq \frac{\alpha\beta}{2}\frac{\ln \mathrm{snr}}{\mathrm{snr}}. \tag{48}$$

5.2.2 Comparison of MMSE and LLSE

To recognize the difference between Eqs. (44) and (47), we calculate the MMSE. We continue to suppose that $\mathrm{snr} = \sigma^2/\theta^2 \geq 1$ (otherwise, σ, θ may be arbitrary positive numbers for the moment). Following [29], given

$$\boldsymbol{Y}^K = \sqrt{\mathrm{snr}}\boldsymbol{H}^K\boldsymbol{X}'^K + \boldsymbol{N}'^K, \tag{49}$$

the MMSE in estimating $\boldsymbol{H}^K\boldsymbol{X}'^K$ is

$$\begin{aligned} \mathrm{mmse}(\mathrm{snr}) &= \mathbb{E}\left\{\left\|\boldsymbol{H}^K\boldsymbol{X}'^K - \boldsymbol{H}^K\widehat{\boldsymbol{X}'^K}\right\|^2\right\} \\ &= \mathrm{tr}\{\boldsymbol{H}^K(\boldsymbol{\Sigma}^{-K} + \mathrm{snr}\boldsymbol{H}^{KT}\boldsymbol{H}^K)^{-1}\boldsymbol{H}^{KT}\}, \end{aligned}$$

where $\widehat{\boldsymbol{X}'^K}$ is the minimum mean-square estimate of $\boldsymbol{X}'^K$, and $\boldsymbol{\Sigma}^K$ is the covariance matrix of $\boldsymbol{X}'^K$. Since $\boldsymbol{X}'^K$ has independent Gaussian components $X'_k \sim \mathcal{N}(0, 1 - \mathrm{snr}^{-1}\rho^{-2k-1})$, $\boldsymbol{\Sigma}^K$ is a $K \times K$ diagonal matrix with entries $\sigma_{kk} = 1 - \mathrm{snr}^{-1}\rho^{-2k-1}$, $k = 0, \ldots, K-1$. A computation yields

$$\mathrm{mmse}(\mathrm{snr}) = \sum_{k=0}^{K-1} \mathrm{snr}^{-1} \left(1 - \mathrm{snr}^{-1}\rho^{-2k-1}\right).$$

When $\alpha\beta$ becomes large, we obtain by transition to a Riemann integral

$$\begin{aligned}
\mathrm{mmse}(\mathrm{snr}) \cdot \delta &= \sum_{k=0}^{K-1} \mathrm{snr}^{-1} \left(1 - \mathrm{snr}^{-1}\rho^{-2k-1}\right) \delta \\
&= \mathrm{snr}^{-1} \int_0^{\infty} \left(1 - \mathrm{snr}^{-1} e^{2x}\right)^+ dx + \epsilon \\
&= \frac{1}{2}\frac{\ln \mathrm{snr}}{\mathrm{snr}} - \frac{1}{2}\frac{1}{\mathrm{snr}}\left(1 - \frac{1}{\mathrm{snr}}\right) + \epsilon,
\end{aligned}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Now, returning to the scaling $\sigma^2 = 1, 0 < \theta^2 \leq 1$, Eq. (48) applies and we obtain, observing $\delta \sim \frac{1}{\alpha\beta}$ as $\alpha\beta \to \infty$,

$$\mathrm{mmse}(\mathrm{snr}) \doteq \mathrm{llse}(\mathrm{snr}) - \frac{\alpha\beta}{2}\frac{1}{\mathrm{snr}}\left(1 - \frac{1}{\mathrm{snr}}\right). \tag{50}$$

By averaging with respect to the DoF $\alpha\beta$, the asymptotic equations (48), (50) turn into true equations and we get for LLSE and MMSE, resp.,

$$\begin{aligned}
\overline{\mathrm{llse}}(\mathrm{snr}) &= \lim_{\alpha\beta\to\infty} \frac{\mathrm{llse}(\mathrm{snr})}{\alpha\beta} = \frac{1}{2}\frac{\ln \mathrm{snr}}{\mathrm{snr}}, \\
\overline{\mathrm{mmse}}(\mathrm{snr}) &= \lim_{\alpha\beta\to\infty} \frac{\mathrm{mmse}(\mathrm{snr})}{\alpha\beta} \\
&= \overline{\mathrm{llse}}(\mathrm{snr}) - \frac{1}{2}\frac{1}{\mathrm{snr}}\left(1 - \frac{1}{\mathrm{snr}}\right).
\end{aligned}$$

In Fig. 11, $\overline{\mathrm{llse}}(\mathrm{snr})$ and $\overline{\mathrm{mmse}}(\mathrm{snr})$ are plotted against $10\log_{10}\mathrm{snr}$ for $\mathrm{snr} \geq 1$.

5.2.3 Application to MIMO Systems

The scaling assumption in Proposition 1 has only been needed to give the RHS of Eq. (47) the meaning of an estimation error. Without this assumption, it always holds that

$$\frac{d}{d\,\mathrm{snr}} C_0(\mathrm{snr}) \doteq \frac{1}{2}\left\{\mathrm{mmse}(\mathrm{snr}) + \frac{\alpha\beta}{2}\frac{1}{\mathrm{snr}}\left(1 - \frac{1}{\mathrm{snr}}\right)\right\}. \tag{51}$$

Since the channel model (43) (or (49), *cf.* [29]) also applies to a MIMO system with K transmit and K receive antennas and additive Gaussian noise, Eq. (51)

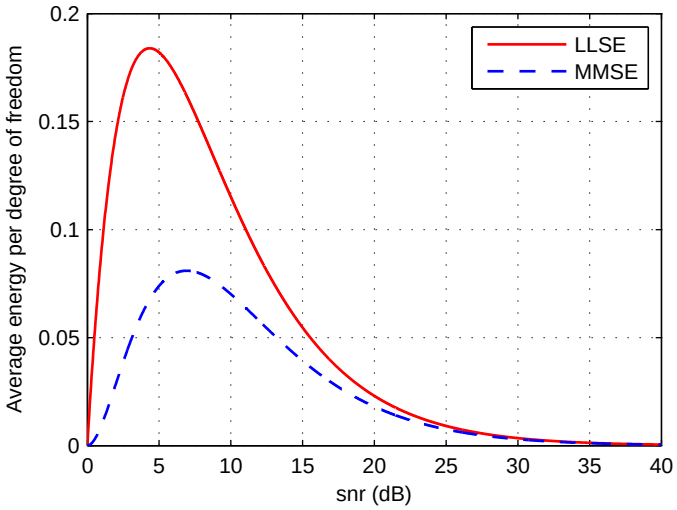

Figure 11: Average energy per degree of freedom of LLSE and MMSE as DoF $\alpha\beta \to \infty$. Scaling $\sigma^2 = 1$, $0 < \theta^2 \leq 1$ of parameters σ^2, θ^2 assumed.

may be used to estimate the capacity of certain MIMO systems. The capacity of a general MIMO system (43) has been computed in [41] and, in case of a diagonal channel matrix, optimal input power allocation policies are given in [42]. However, in contrast to [29], [41], [42], and [43], in our setting the probability distribution of optimal input signals depends on snr. Now, Eq. (51) tells us that, when $\alpha\beta$ is sufficiently large, with growing snr [thus, by Eq. (24), $K \doteq \alpha\beta \ln \sqrt{\mathsf{snr}} \to \infty$] the increase of channel capacity is significantly higher than anticipated by the I-MMSE relationship (44), at least in the lower snr region (*cf.* Fig. 11). The import of this observation is that high-dimensional MIMO systems might perform significantly better than hitherto assumed.

6 Conclusion

In this paper, we considered the heat channel, in its most basic form a linear time-varying analog filter with subsequent AWGN. The representation of the filter by means of its orthonormal eigenfunctions (dilated Hermite functions) not only leads to a diagonalization of the filter but also allows for a precise definition of measurement of the noisy filter output in terms of optimal detection. This leads to a discretization of the original continuous-time heat channel and results in its final form, an infinite-dimensional vector Gaussian channel (where the measurement error is incorporated into the noise variances). Under a certain growth condition on the average input energy, a further reduction to a finite set of parallel Gaussian channels is possible resulting in an closed-form expression for channel capacity in terms of the input energy. We then pro-

posed a method for capacity achieving signaling. The capacity formula is an asymptotic equation; time averaging yields an equation and capacity (in bits per second) as a function of average input power thus allowing a comparison with the capacity (or spectral efficiency) of the bandlimited Gaussian channel, and of the Gallager channel with Gaussian filter. Surprisingly, the spectral efficiency of the Gallager channel (with Gaussian filter) is, in general, inferior to that of the heat channel. These findings give evidence that the classical capacity result [7, Theorem 8.5.1] may lead to an overly conservative assessment of the attainable information throughput of communication systems (*e.g.*, in fiber optics in the presence of dispersion and amplifier noise). On the other hand, we found that the water-filling characterization of the capacity of the heat channel in the time-frequency plane and that of the Gallager channel (with Gaussian filter) in the frequency domain compare well. We then computed the rate distortion function for white Gaussian noise response of the filter. The Lambert W function proved to be an indispensable tool for the statement of closed-form channel capacity/rate distortion formulas. Since the Lambert W function is not an elementary function, it was noticed that a simpler capacity formula based on the measured filter output signal is possible. The new setting was put into the framework of estimation theory resulting in a "C-LLSE formula" connecting the capacity of the heat channel with an estimation error for the output. Thus, in a sense, that particular formula can be seen as a parallel to the general I-MMSE relationship discovered in [29] (where SNR-dependent inputs are precluded). The interpretation of our results in context of MIMO systems indicate that high-dimensional MIMO systems might perform significantly better than assumed.

Appendix

A Wigner-Ville spectrum of Filter Response on White Gaussian Noise

We model white Gaussian noise of two-sided noise PSD $\sigma^2 \in (0, \infty)$ by a sequence of stochastic processes $\{\boldsymbol{U}^K(t), t \in \mathbb{R}\}$, $K = 1, 2, \ldots$, given by their respective Karhunen-Loève expansion

$$\boldsymbol{U}^K(t) = \sum_{k=0}^{K-1} U_k \, (D_\gamma \psi_k)(t), \; t \in \mathbb{R},$$

where $U_0, \ldots, U_{K-1}$ are i.i.d. Gaussian random variables $\sim \mathcal{N}(0, \sigma^2)$. For any $K = 1, 2, \ldots$, let the process $\{\boldsymbol{U}^K(t), t \in \mathbb{R}\}$ be the input to filter $\boldsymbol{P}_\delta^{(\gamma)}$. By means of representation (8) it is seen that the corresponding filter output tends as $K \to \infty$ to the stochastic process

$$X(t) = \sum_{k=0}^{\infty} \rho^{k+\frac{1}{2}} U_k \, (D_\gamma \psi_k)(t), \; t \in \mathbb{R}. \tag{52}$$

We interprete $\{X(t), t \in \mathbb{R}\}$ as filter response on white Gaussian noise.

Since any realization $x(t)$ of $\{X(t)\}$ is almost surely in $L^2(\mathbb{R})$, the Wigner distribution [44]

$$(Wx)(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t'} x\left(t+\frac{t'}{2}\right)\overline{x\left(t-\frac{t'}{2}\right)}\,dt'$$

may be computed. By taking the ensemble average, we obtain the WVS [27] of process $\{X(t)\}$,

$$\begin{aligned}\mathbf{\Phi}(t,\omega) &= \mathbb{E}\{(WX)(t,\omega)\}\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t'} r\left(t+\frac{t'}{2}, t-\frac{t'}{2}\right)\,dt', \qquad (53)\end{aligned}$$

where $r(t_1,t_2) = \mathbb{E}\{X(t_1)\overline{X(t_2)}\}$ is the autocorrelation function. The kernel of operator $\boldsymbol{P}_\delta^{(\gamma)}$ has for arbitrary parameters $\gamma > 0, \delta > 0$ two alternative representations (following from a slight generalization of Mehler's formula, see [14]),

$$\begin{aligned}&P_\delta^{(\gamma)}(x,y)\\ &= \sum_{k=0}^{\infty}\rho^{k+\frac{1}{2}}(D_\gamma\psi_k)(x)(D_\gamma\psi_k)(y)\\ &= \frac{1}{\gamma\sqrt{2\pi\sinh\delta}}\exp\left\{-\frac{1}{4\gamma^2}\left[\coth\left(\frac{\delta}{2}\right)(x-y)^2+\tanh\left(\frac{\delta}{2}\right)(x+y)^2\right]\right\}.\end{aligned}$$

We infer by the first representation that $r(t_1,t_2) = \sigma^2 P_{2\delta}^{(\gamma)}(t_1,t_2)$. Then, by means of the second representation, the integral in (53) is readily evaluated and we obtain for the WVS

$$\mathbf{\Phi}(t,\omega) = \frac{\sigma^2}{2\pi}\cdot\frac{1}{\cosh\delta}\exp\left(-\frac{t^2}{\alpha^2}-\frac{\omega^2}{\beta^2}\right).$$

B Proofs of Theorems 2 and 4

For any bounded operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ the Weyl symbol $\sigma_A(x,\xi)$ is defined by

$$(Af)(x) = \frac{1}{2\pi}\iint_{\mathbb{R}^2}\sigma_A\left(\frac{x+y}{2},\xi\right)e^{i(x-y)\xi}f(y)\,dy\,d\xi, \qquad (54)$$

see, *e.g.*, [44], [25], [24]; the linear map $A \mapsto \sigma_A(x,\xi)$ is called Weyl correspondence. For example, the operator $A = \boldsymbol{P}_{2\delta}^{(\gamma)}$ has the Weyl symbol [14]

$$\begin{aligned}\sigma_A(x,\xi) &= \frac{1}{\cosh\delta}e^{-(\tanh\delta)(\gamma^{-2}x^2+\gamma^2\xi^2)} \qquad (55)\\ &= \frac{1}{\cosh\delta}\exp\left(-\frac{x^2}{\alpha^2}-\frac{\xi^2}{\beta^2}\right).\end{aligned}$$

In the rest of this appendix, A will always stand for operator $\boldsymbol{P}_{2\delta}^{(\gamma)}$, and λ_k, $k = 0, 1, \ldots$, for its eigenvalues $\rho^{2k+1} \in (0, 1)$. The proof of the subsequent Theorem 6 follows the argument in [24] (*cf.* also [45]); the Szegö theorems in [24], [45] are inadequate for our purposes, though.

Lemma 1 *For any polynomial* $G_N(x, z) = \sum_{n=1}^{N} c_n(x) z^n$ *with bounded variable coefficients* $c_n(x) \in \mathbb{R}$, $x \in (1, \infty)$, *it holds*

$$\sum_{k=0}^{\infty} G_N(\alpha\beta, \lambda_k) \doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} G_N\left(\alpha\beta, \sigma_A(x, \xi)\right) \, dx \, d\xi. \tag{56}$$

Proof: First, by (8), for any $f \in L^2(\mathbb{R})$ it holds that

$$G_N(\alpha\beta, A) f = \sum_{k=0}^{\infty} G_N(\alpha\beta, \lambda_k) \langle f, D_\gamma \psi_k \rangle D_\gamma \psi_k.$$

Hence, operator $B = G_N(\alpha\beta, A)$ has the trace

$$\operatorname{tr} B = \sum_{k=0}^{\infty} G_N(\alpha\beta, \lambda_k). \tag{57}$$

Secondly, we use the key observation [24, trace rule (0.4)] to obtain (here and thereafter, double integrals extend over $\mathbb{R}^2$)

$$\operatorname{tr} B = \frac{1}{2\pi} \iint \sigma_B(x, \xi) \, dx \, d\xi,$$

where $\sigma_B(x, \xi)$ is the Weyl symbol of operator B. By linearity of the Weyl correspondence, the Weyl symbol of B has the expansion

$$\sigma_B(x, \xi) = \sum_{n=1}^{N} c_n(\alpha\beta) \sigma_{A^n}(x, \xi).$$

Since for any $\gamma > 0$ held constant the family of operators $\{\boldsymbol{P}_{\delta}^{(\gamma)}; \delta > 0\}$ forms a semigroup with respect to δ (see [14]), it follows that $A^n = \boldsymbol{P}_{2n\delta}^{(\gamma)}$. In Eq. (55), replace operator A by A^n and δ by $n\delta$. Because of $\tanh(n\delta) = (n \tanh \delta)(1 + o(1))$ we then obtain

$$\begin{aligned} \sigma_{A^n}(x, \xi) &= \frac{1}{\cosh(n\delta)} e^{-\tanh(n\delta)(\gamma^{-2}x^2 + \gamma^2\xi^2)} \\ &= (1 + o(1)) \left(\sigma_A(x, \xi)\right)^n \exp\left[-o(1)\left(\frac{x^2}{\alpha^2} + \frac{\xi^2}{\beta^2}\right)\right], \end{aligned}$$

where the Landau symbol $o(1)$ stands for various quantities vanishing as $\delta \to 0$

(or $\alpha\beta \to \infty$). We now estimate

$$\begin{aligned} \operatorname{tr} B &= \frac{1}{2\pi} \iint \sigma_B(x,\xi)\, dx\, d\xi \\ &= \left[\frac{1}{2\pi} \iint G_N\left(\alpha\beta, \sigma_A(\alpha x, \beta\xi)\right)\, dx\, d\xi + \epsilon \right] \alpha\beta \\ &= \frac{1}{2\pi} \iint G_N(\alpha\beta, \sigma_A(x,\xi))\, dx\, d\xi + \epsilon\, \alpha\beta, \end{aligned} \tag{58}$$

where $\epsilon \to 0$ as $\alpha\beta \to \infty$. Eq. (58) in combination with Eq. (57) concludes the proof. $\square$

Theorem 6 (Szegö Theorem) *Let $g : [0,\Delta] \to \mathbb{R}$, $\Delta \in (0,\infty)$, be a continuous function such that $\lim_{x\to 0+} g(x)/x$ exists. For any functions $a, b : (1,\infty) \to \mathbb{R}$, where $a(x)$ is bounded and $b(x) \in [0,\Delta]$, define the function $G(x,z) = a(x)g(b(x)z)$, $(x,z) \in (1,\infty)\times[0,1]$. Then it holds*

$$\sum_{k=0}^{\infty} G(\alpha\beta, \lambda_k) \doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} G\left(\alpha\beta, \sigma_A(x,\xi)\right)\, dx\, d\xi. \tag{59}$$

Proof: The function $f(x) = g(x)/x$, $x \in (0,\Delta]$, has a continuous extension $F(x)$ onto the compact interval $[0,\Delta]$. By virtue of the Weiertstrass approximation theorem, for any $n \in \mathbb{N}$ there exists a polynomial $F_{N_n-1}(x)$ of some degree $N_n - 1$ such that $|F(x) - F_{N_n-1}(x)| \le \epsilon_n = \frac{1}{n}$ for all $x \in [0,\Delta]$. Consequently, the polynomial $g_{N_n}(x) = xF_{N_n-1}(x)$ of degree N_n satisfies the inequality

$$|g(x) - g_{N_n}(x)| \le \epsilon_n x,\ x \in [0,\Delta]. \tag{60}$$

Define the polynomial with variable coefficients $G_{N_n}(x,z) = a(x)g_{N_n}(b(x)z)$. We now show that

$$(\alpha\beta)^{-1} \sum_{k=0}^{\infty} G_{N_n}(\alpha\beta, \lambda_k) \to (\alpha\beta)^{-1} \sum_{k=0}^{\infty} G(\alpha\beta, \lambda_k) \tag{61}$$

and

$$\begin{aligned} &\frac{(\alpha\beta)^{-1}}{2\pi} \iint_{\mathbb{R}^2} G_{N_n}\left(\alpha\beta, \sigma_A(x,\xi)\right)\, dx\, d\xi \\ &\quad\to \frac{(\alpha\beta)^{-1}}{2\pi} \iint_{\mathbb{R}^2} G\left(\alpha\beta, \sigma_A(x,\xi)\right)\, dx\, d\xi, \end{aligned} \tag{62}$$

as $n \to \infty$, uniformly for $\alpha\beta \in (1,\infty)$.

Proof of (61): By Ineq. (60) we get

$$\begin{aligned} \Big|\sum_{k=0}^{\infty} G(\alpha\beta,\lambda_k) - \sum_{k=0}^{\infty} G_{N_n}(\alpha\beta,\lambda_k)\Big| &\le \sum_{k=0}^{\infty} |G(\alpha\beta,\lambda_k) - G_{N_n}(\alpha\beta,\lambda_k)| \\ &\le M\epsilon_n \Delta \sum_{k=0}^{\infty} \lambda_k, \end{aligned}$$

where $M = \sup\{|a(x)|; x > 1\} < \infty$ and $\sum_{k=0}^{\infty} \lambda_k = \rho/(1-\rho^2) \leq \alpha\beta/2$ for $\alpha\beta > 1$. After devision of the last inequality by $\alpha\beta$, convergence in (61) follows as claimed.

Proof of (62): Similarly,

$$\begin{aligned} &\left| \iint G\left(\alpha\beta, \sigma_A(x,\xi)\right)\, dx\, d\xi - \iint G_{N_n}\left(\alpha\beta, \sigma_A(x,\xi)\right)\, dx\, d\xi \right| \\ &\quad \leq \iint \left| G\left(\alpha\beta, \sigma_A(x,\xi)\right) - G_{N_n}\left(\alpha\beta, \sigma_A(x,\xi)\right) \right| dx\, d\xi \\ &\quad \leq M\epsilon_n \Delta \iint \sigma_A(x,\xi)\, dx\, d\xi. \end{aligned}$$

Since $(2\pi)^{-1} \iint \sigma_A(x,\xi)\, dx\, d\xi = \rho/(1-\rho^2)$, we arrive at the same conclusion as before.

Now, choose any number $n \in \mathbb{N}$, substitute function G in Eq. (59) by the polynomial G_{N_n} and devide both sides of that equation by $\alpha\beta$. Let $n \to \infty$. Then, by reason of Lemma 1 and uniform convergence in (61) and (62) with respect to $\alpha\beta \in (1, \infty)$, the theorem follows. □

Proof of Theorem 2: Define

$$\ln_+ x = \begin{cases} \max\{0, \ln x\} & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Because of Eq. (17) we have, recalling that σ^2 is dependent on $\alpha\beta$,

$$\begin{aligned} C(S) &= \sum_{k=0}^{\infty} \frac{1}{2} \ln_+ \left(\frac{\sigma^2(\alpha\beta)}{\theta^2} \lambda_k \right) \\ &= \sum_{k=0}^{\infty} a(\alpha\beta) g(b(\alpha\beta)\lambda_k), \end{aligned}$$

where $a(\alpha\beta) = 1$, $b(\alpha\beta) = \sigma^2(\alpha\beta)/\theta^2$, $g(x) = \frac{1}{2}\ln_+ x, x \in [0, \Delta]$, and Δ is chosen so that $b(\alpha\beta) \leq \Delta < \infty$ when $\alpha\beta$ is large enough (the latter choice is possible since $\sigma^2(\alpha\beta)$ is finitely upper bounded as $\alpha\beta \to \infty$). Without loss of generality, we assume $b(\alpha\beta) \in [0, \Delta]$ for all $\alpha\beta \in (1, \infty)$. Then, by Theorem 6 it follows that

$$\begin{aligned} C(S) &\doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} \frac{1}{2} \ln_+ \left(\frac{\sigma^2(\alpha\beta)}{\theta^2} \sigma_A(x,\xi) \right) dx\, d\xi \\ &= \frac{1}{2\pi} \iint \frac{1}{2} \ln \left[1 + \frac{\left(\frac{\sigma^2(\alpha\beta)}{2\pi} - N(x,\xi) \right)^+}{N(x,\xi)} \right] dx\, d\xi, \end{aligned}$$

where $N(x,\xi) = \frac{\theta^2}{2\pi}(\sigma_A(x,\xi))^{-1}$. Next, rewrite Eq. (14) as

$$S = \sum_{k=0}^{\infty} \sigma^2(\alpha\beta) \left(1 - \frac{1}{\frac{\sigma^2(\alpha\beta)}{\theta^2}\lambda_k} \right)^+.$$

Set $a(\alpha\beta) = \sigma^2(\alpha\beta)$, $b(\alpha\beta) = \sigma^2(\alpha\beta)/\theta^2$ and define

$$g(x) = \begin{cases} \left(1 - \frac{1}{x}\right)^+ & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Without loss of generality, we assume that $a(\alpha\beta)$ is bounded for all $\alpha\beta \in (1, \infty)$. So, $b(\alpha\beta) \in [0, \Delta]$ where $\Delta = \sup\{a(\alpha\beta)/\theta^2; \alpha\beta > 1\} < \infty$. Then, by Theorem 6 it follows that

$$\begin{aligned} S &\doteq \frac{1}{2\pi} \iint \sigma^2(\alpha\beta) \left(1 - \frac{1}{\frac{\sigma^2(\alpha\beta)}{\theta^2}\sigma_A(x,\xi)}\right)^+ dx\, d\xi \\ &= \iint \left(\frac{\sigma^2(\alpha\beta)}{2\pi} - N(x,\xi)\right)^+ dx\, d\xi. \end{aligned}$$

Finally, replacement of $\frac{\sigma^2(\alpha\beta)}{2\pi}$ by the parameter ν completes the proof. □

Proof of Theorem 4: For any $\theta \in (0, \sigma]$ held constant define the distortion D by Eq. (35) or, equivalently, by

$$D = \sum_{k=0}^{\infty} \theta^2 \min\left\{1, \frac{\sigma^2}{\theta^2}\lambda_k\right\}.$$

Since $D = \sum_{k=0}^{\infty} a(\alpha\beta) g(b(\alpha\beta)\lambda_k)$, where $a(\alpha\beta) = \theta^2$, $b(\alpha\beta) = \sigma^2/\theta^2$, $g(x) = \min\{1, x\}$ for $x \in [0, \Delta]$, $\Delta = \sigma^2/\theta^2$, it follows by Theorem 6 that

$$\begin{aligned} D &\doteq \frac{1}{2\pi} \iint \theta^2 \min\left\{1, \frac{\sigma^2}{\theta^2}\sigma_A(x,\xi)\right\} dx\, d\xi \qquad (63) \\ &= \iint \min\left\{\frac{\theta^2}{2\pi}, \mathbf{\Phi}(t,\omega)\right\} dt\, d\omega, \end{aligned}$$

where $\mathbf{\Phi}(t,\omega) = \frac{\sigma^2}{2\pi}\sigma_A(t,\omega)$ is the WVS (12). Next, rewrite Eq. (38) as

$$R = \sum_{k=0}^{\infty} \frac{1}{2} \ln_+\left(\frac{\sigma^2}{\theta^2}\lambda_k\right).$$

Taking $a(\alpha\beta) = 1$, $b(\alpha\beta) = \sigma^2/\theta^2$, $g(x) = \frac{1}{2}\ln_+ x, x \in [0, \Delta]$, Δ chosen as before, we infer by Theorem 6 that

$$\begin{aligned} R &\doteq \frac{1}{2\pi} \iint_{\mathbb{R}^2} \frac{1}{2} \ln_+\left(\frac{\sigma^2}{\theta^2}\sigma_A(x,\xi)\right) dx\, d\xi \qquad (64) \\ &= \frac{1}{2\pi} \iint \frac{1}{2} \ln_+\left[\frac{\mathbf{\Phi}(t,\omega)}{\frac{\theta^2}{2\pi}}\right] dt\, d\omega. \end{aligned}$$

Finally, replacement of $\frac{\theta^2}{2\pi}$ by the parameter λ completes the proof. □

Acknowledgment

We are grateful to the reviewers and the Associate Editor, Dongning Guo, for their helpful comments, remarks, and suggestions. We want to thank Peter Jung for interesting discussions about Szegö theorems.

References

[1] J. Fourier, *The Analytical Theory of Heat*. Mineola, NY: Dover Publ., 2003. (engl. transl. of Fourier's 1822 book)

[2] A. N. Tikhonov, V. Y. Arsenin, and F. John (Transl.), *Solutions of ill-posed problems*, Washington, DC: Winston, 1977.

[3] D. D. Falconer, "History of equalization 1860–1980," *IEEE Commun. Mag.*, vol. 49, pp. 42–50, 2011.

[4] G. P. Agrawal, *Fiber-optic communication systems*. 3rd ed. New York, NY: Wiley, 2002.

[5] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, pp. 662–701, 2010.

[6] P. J. Winzer and R.-J. Essiambre, "Advanced optical modulation formats," in *Optical Fiber Telecommunications V B*, I. P. Kaminov, T. Li, and A. E. Willner, Eds. San Diego, CA: Academic Press, pp. 23–94, 2008.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[8] S. M. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, pp. 207–223, 2012.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley, 2006.

[11] C. R. Baker, "Information capacity of the stationary Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 37, pp. 1314–1326, 1991.

[12] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[13] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inf. Theory*, vol. 34, pp. 605–612, 1988.

[14] E. Hammerich, "A sampling theorem for time-frequency localized signals," *Sampl. Theory Signal Image Process.*, vol. 3, pp. 45–81, 2004.

[15] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.* (London), vol. 93 (III), pp. 429–457, 1946.

[16] E. Getzler, "A short proof of the local Atiyah-Singer index theorem," *Topology*, vol. 25, pp. 111–117, 1985.

[17] J. L. Holsinger, *Digital communication over fixed time-continuous channels with memory, with special application to telephone channels*. Tech. Rept. 430, MIT Research Lab. of Electronics, Cambridge, MA, 1964.

[18] P. M. Ebert, *Error bounds for parallel communication channels*. Tech. Rept. 448, MIT Research Lab. of Electronics, Cambridge, MA, 1966.

[19] M. Katz, W. L. Murdock, and G. Szegö, "On the eigenvalues of certain Hermitian forms," *J. Rat. Mech. and Anal.*, vol. 2, pp. 767-800, 1953.

[20] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. Berkeley, CA: Univ. of Calif. Press, 1958.

[21] P. Jung, "On the Szegö-asymptotics for doubly-dispersive Gaussian channels," *Proc. IEEE Int. Symp. Information Theory*, St. Petersburg, Russia, 2011, pp. 2852–2856.

[22] B. Farrell and T. Strohmer, "Eigenvalue estimates and mutual information for the linear time-varying channel," *IEEE Trans. Inf. Theory*, vol. 57, pp. 5710–5718, 2011.

[23] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 367–395, 2010.

[24] A. J. E. M. Janssen and S. Zelditch, "Szegö limit theorems for the harmonic oscillator," *Trans. Amer. Math. Soc.*, vol. 280, pp. 563–587, 1983.

[25] W. Kozek and F. Hlawatsch, "Time-frequency representation of linear time-varying systems using the Weyl symbol," in *Proc. IEE Sixth Int. Conf. on Digital Signal Process. in Commun.,* Loughborough, UK, pp. 25–30, 1991.

[26] F. Hlawatsch and W. Kozek, "Second-order time-frequency synthesis of nonstationary random processes," *IEEE Trans. Inf. Theory*, vol. 41, pp. 255–267, 1995.

[27] P. Flandrin and W. Martin, "The Wigner-Ville spectrum of nonstationary random signals," in *The Wigner Distribution*, W. Mecklenbräuker and F. Hlawatsch, Eds. Amsterdam: Elsevier, 1997, pp. 211–267.

[28] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[29] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, 2005.

[30] Y. Wu, D. Guo, and S. Verdú, "Derivative of mutual information at zero SNR: the Gaussian noise case," *IEEE Trans. Inf. Theory*, vol. 57, pp. 7307–7312, 2011.

[31] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York, NY: Dover Publ., 1972.

[32] D. O. North, *Analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems*. Rept. PTR-6C, RCA Labs., Princeton, NJ, 1943.

[33] G. L. Turin, "An introduction to matched filters," *IRE Trans. Inf. Theory*, vol. 6, pp. 311–329, 1960.

[34] E. Hammerich, "On the heat channel and its capacity," *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, 2009, pp. 1809–1813.

[35] H. Gross, *Handbook of Optical Systems, Vol. 1.* Weinheim: Wiley-VCH, 2005.

[36] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[37] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Computational Math.*, vol. 5, pp. 329–359, 1996.

[38] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, pp. 622–640, 1999.

[39] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[40] A. N. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, pp. 102–108, 1956.

[41] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[42] A. L. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3033–3051, 2006.

[43] D. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, pp. 453–470, 2007.

[44] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Boston: Birkhäuser, 2001.

[45] H. G. Feichtinger and K. Nowak, "A Szegö-type theorem for Gabor-Toeplitz localization operators," *Michigan Math. J.*, vol. 49, pp. 13–21, 2001.